$$c_2(\boldsymbol{u}) = 1 - u_1^2 - u_2^2 = 0$$

- This situation is depicted in Fig 4.1

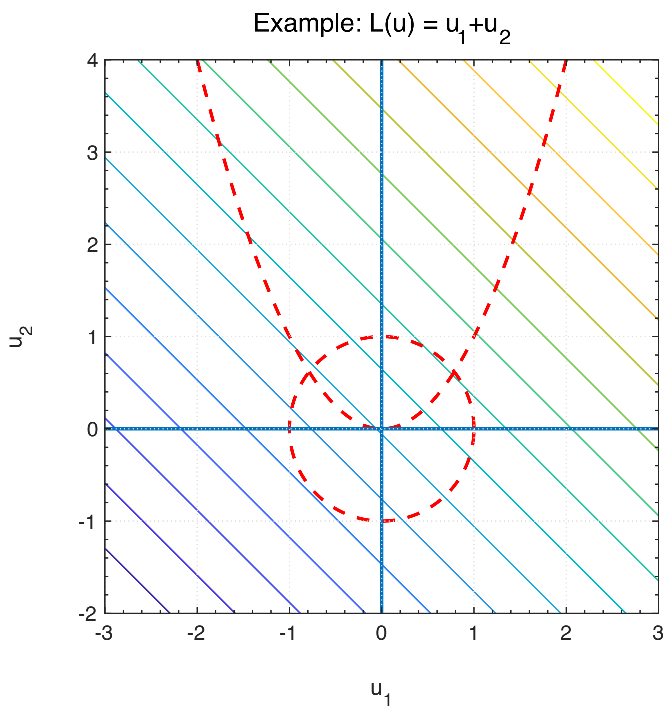

Figure 4.1: Constrained Optimization Example

- The straight-line contours depict constant values of the objective function $L(\boldsymbol{u})$, with smaller values shown by the darker colors.
- The red dashed curves show the two equality constraints.
- It is apparent, by inspection, that satisfaction of the equality constraints occurs at two distinct points, which may be found by solving the constraint equations simultaneously
  - Adding $c_1(\boldsymbol{u}) + c_2(\boldsymbol{u})$ gives the polynomial,

$$u_2^2 + u_2 - 1 = 0$$

which yields the solutions

$$u_2 = -1.618,\ 0.618$$

- Since only one of these can give a point on the unit circle, we subtitute $u_2 = 0.618$ into $u_1^2 + u_2^2 = 1$ and solve for u_1 which gives

$$u_1 = \pm\sqrt{1 - u_2^2} = \pm 0.7862$$

- Hence both constraints are simultaneously met at points: $u_1 = \pm 0.7862,\ u_2 = 0.618$

- Therefore, the solution to this equality constraint problem reduces to selecting which of the two points gives the lower value of $L(\boldsymbol{u})$ (the left-hand point, in this case).

- This example demonstrates the important relationship between numbers of decision variables and the number of constraints.
  - In this case, practically all degrees of freedom are used satisfying constraints!

- This chapter will introduce techniques for solving the equality constraint problem mathematically.

4.2: Equality Constraints: Two-Parameter Problem

- Consider the following parameter optimization problem:
  - Find the parameter vector $\boldsymbol{u} = \begin{bmatrix} u_1 & u_2 \end{bmatrix}^T$ that minimizes $L(\boldsymbol{u})$ subject to the equality constraint $c(\boldsymbol{u}) = \gamma$
  - QUESTION: why must there be at least *two* parameters in this problem?
- From the previous chapter, we know that the change in $L(\boldsymbol{u})$ caused by changes in the parameters u_1 and u_2 in a neighborhood of the optimal solution is given (approximately) by:

$$\triangle L \approx \left.\frac{\partial L}{\partial u_1}\right|_* \triangle u_1 + \left.\frac{\partial L}{\partial u_2}\right|_* \triangle u_2 + \frac{1}{2}\triangle \boldsymbol{u}^T \left[\frac{\partial^2 L}{\partial \mathbf{u}^2}\right]_* \triangle \boldsymbol{u}$$

  - But as we change the point $(u_1,\ u_2)$ we must ensure the constraint remains satisfied
  - To first order, this means,

$$\triangle c = \left.\frac{\partial c}{\partial u_1}\right|_* \triangle u_1 + \left.\frac{\partial c}{\partial u_2}\right|_* \triangle u_2 = 0$$

  - So $\triangle u_1$ (or $\triangle u_2$) is *not* arbitrary; it depends on $\triangle u_2$ (or $\triangle u_1$) according to the following relationship:

$$\triangle u_1 = -\left\{\frac{\partial c/\partial u_2|_*}{\partial c/\partial u_1|_*}\right\} \triangle u_2$$

$\Rightarrow$

$$\triangle L \approx \left\{\frac{\partial L}{\partial u_2} - \frac{\partial L}{\partial u_1}\left[\frac{\partial c/\partial u_2}{\partial c/\partial u_1}\right]\right\}_* \triangle u_2$$

$$+\frac{1}{2}\left\{\frac{\partial^2 L}{\partial u_2^2} - 2\frac{\partial^2 L}{\partial u_1 \partial u_2}\left[\frac{\partial c/\partial u_2}{\partial c/\partial u_1}\right] + \frac{\partial^2 L}{\partial u_1^2}\left[\frac{\partial c/\partial u_2}{\partial c/\partial u_1}\right]^2\right\}_* \triangle u_2^2$$

– But since $\triangle u_2$ *is* arbitrary, how do we find the solution?

◦ Coefficient of first-order term must be zero

◦ Coefficient of second-order term must be greater than or equal to zero

- Hence we can solve this problem by solving

$$\left.\frac{\partial L}{\partial u_2}\right|_* - \frac{\partial L}{\partial u_1}\left[\frac{\partial c/\partial u_2}{\partial c/\partial u_1}\right] = 0$$

– But this is only one equation in two unknowns; where do we get the rest of the information required to solve this problem?

◦ From the *constraint*!

$$c(u_1,\, u_2) = \gamma$$

- Although the approach outlined above is straightforward for the 2-parameter/1-constraint problem, it becomes difficult to implement as the dimensions of the problem increase

- For this reason, it would be nice to develop an alternative approach that can be extended to more difficult problems

- Is this possible?

$$\Rightarrow \text{YES!}$$

4.3: Lagrange Multipliers: Two-Parameter Problem

- An alternative to the approach considered in the previous section will now be developed.
- The basic idea is to modify the optimization cost function to incorporate equality constraints *directly*.
- Consider the following augmented form of the cost function $L(u_1,\, u_2)$:

$$\bar{L} = L(u_1,\, u_2) + \lambda\left\{c(u_1,\, u_2) - \gamma\right\}$$

  where λ is a constant we are free to select

  - Since $c(u_1,\, u_2) - \gamma = 0$, this new cost function will be minimized at precisely the same points as $L(u_1,\, u_2)$
  - Therefore, a NECESSARY CONDITION for a local minimum of $L(u_1,\, u_2)$ is given by

$$\triangle\bar{L} = \triangle L + \lambda\triangle c = 0$$

    which may be expressed as,

$$\left\{\frac{\partial L}{\partial u_1} + \lambda\frac{\partial c}{\partial u_1}\right\}\triangle u_1 + \left\{\frac{\partial L}{\partial u_2} + \lambda\frac{\partial c}{\partial u_2}\right\}\triangle u_2 = 0$$

  - But because of the constraint, $\triangle u_1$ and $\triangle u_2$ are not independent
  - So now it is conceivable that this result could be true even if the two coefficients are not zero
  - However, λ is free to be chosen
  - Let

$$\lambda = \frac{-\partial L/\partial u_1}{\partial c/\partial u_1}$$

    then we may write

$$\frac{\partial L}{\partial u_2} + \lambda\frac{\partial c}{\partial u_2} = 0$$

and

$$c(u_1,\ u_2) = 0$$

– Thus we have 3 equations and 3 unknowns

$\Rightarrow$ Can find a unique solution

- Comments:

– Is this a new result? No!

$\Rightarrow$ By substituting the expression for λ into the second equation, you get the *same result* as developed previously!

– So, *why do it*?

– It turns out these equations are precisely what I would have obtained if I had assumed $\triangle u_1$ and $\triangle u_2$ were *independent*

$$\frac{\partial L}{\partial u_1} + \lambda \frac{\partial c}{\partial u_1} = 0$$

$$\frac{\partial L}{\partial u_2} + \lambda \frac{\partial c}{\partial u_2} = 0$$

– So the use of Lagrange multipliers allows us to develop the necessary conditions for a *constrained* minimum using standard *unconstrained* parameter optimization techniques

$\Rightarrow$ A great simplification for complicated problems!

Example 4.1

- Determine the rectangle of maximum area that can be inscribed inside a circle of radius R.

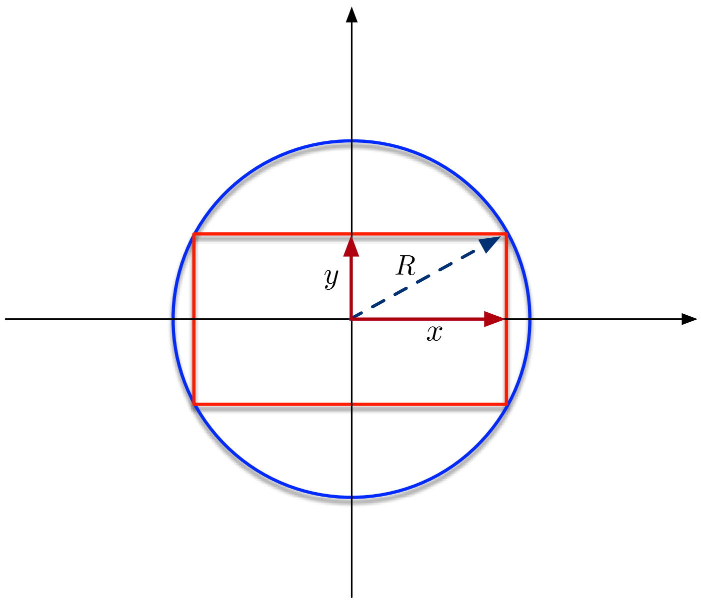

Figure 4.2: Example - Constrained Optimization

- Generate objective function, $L(x,\,y)$:

$$A = 4xy \quad \Rightarrow \quad L = -4xy$$

- Formulate constraint equation, $c(x,\,y)$:

$$c(x,\,y) = x^2 + y^2 = R^2$$

- Calculate first-order conditions:

$$\frac{\partial L}{\partial x} = -4y$$

$$\frac{\partial c}{\partial x} = 2x$$

$$\frac{\partial L}{\partial y} = -4x$$

$$\frac{\partial c}{\partial y} = 2y$$

- Compute the Lagrange multiplier:

$$\lambda = \frac{-\partial L/\partial x}{\partial c/\partial y} = \frac{-(-4y)}{2x} = 2\frac{y}{x}$$

then,

$$\begin{aligned}\frac{\partial L}{\partial y} + \lambda\frac{\partial c}{\partial y} &= -4x + 2\left(\frac{y}{x}\right)(2y) \\ &= -4x + 4\left(\frac{y^2}{x}\right) = 0 \\ &= -4x^2 + 4y^2 = 0\end{aligned}$$

which gives,

$$x^2 = y^2$$

- Combining with the constraint equation,

$$\begin{aligned} x^2 + y^2 + x^2 - y^2 &= R^2 \\ 2x^2 &= R^2 \\ x^2 &= \frac{R^2}{2}\end{aligned}$$

- Thus we arrive at the solution,

$$x^* = \frac{R}{\sqrt{2}}$$

$$y^* = \frac{R}{\sqrt{2}}$$

with maximizing area

$$A^* = 2R^2$$

- The resulting Lagrange multiplier is $\lambda = 2 \quad \Rightarrow$ Is this important?

4.4: Lagrange Multipliers: Multi-Parameter Problem

- The same approach used for the two-parameter problem can also be applied to the multi-parameter problem
- But now for convenience, we'll adopt vector notation:
  - Cost Function: $L(\boldsymbol{x}, \boldsymbol{u})$
  - Constraints: $\boldsymbol{c}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{\gamma}$, where $\boldsymbol{c}$ is dimension $n \times 1$ (i.e., we have n constraint equations)
  - Here we've introduced some new notation:
    - By convention, $\boldsymbol{x}$ will denote the set of *dependent* variables and $\boldsymbol{u}$ will denote the set of *independent* variables in our problem
    - DEPENDENT variables are defined as those whose degrees of freedom are deployed to satisfy constraints
    - INDEPENDENT variables are those remaining to solve the minimization
  - How many dependent variables appear in the constrained problem? $\Rightarrow n$ (why?)

$$\Rightarrow \quad \boldsymbol{x} \equiv (n \times 1) \qquad \boldsymbol{u} \equiv (m \times 1)$$

  where now the length of our overall parameter vector is $m + n$

- Employing the METHOD OF FIRST VARIATIONS, we write,

$$\triangle L \approx \frac{\partial L}{\partial \boldsymbol{x}} \triangle \boldsymbol{x} + \frac{\partial L}{\partial \boldsymbol{u}} \triangle \boldsymbol{u}$$

$$\triangle \boldsymbol{c} \approx \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}} \triangle \boldsymbol{x} + \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{u}} \triangle \boldsymbol{u}$$

$$\Rightarrow \triangle \boldsymbol{x} = -\left\{\frac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}}\right\}^{-1} \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{u}} \triangle \boldsymbol{u}$$

– NOTE: $\boldsymbol{x}$ must be selected so that matrix $\partial \boldsymbol{c}/\partial \boldsymbol{x}$ is *nonsingular*; for well-posed systems, this can always be done.

– Using this expression for $\triangle \boldsymbol{x}$, the necessary conditions for a minimum can now be written:

$$\frac{\partial L}{\partial \boldsymbol{u}} - \frac{\partial L}{\partial \boldsymbol{x}} \left\{ \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}} \right\}^{-1} \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{u}} = 0$$

$$\boldsymbol{c}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{\gamma}$$

giving $(m+n)$ equations in $(m+n)$ unknowns.

- Using the METHOD OF LAGRANGE MULTIPLIERS, we first construct the augmented cost function,

$$\bar{L} = L(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{\lambda}^T \{\boldsymbol{c}(\boldsymbol{x}, \boldsymbol{u}) - \boldsymbol{\gamma}\}$$

where $\boldsymbol{\lambda}$ is now a $(1 \times n)$ vector of parameters that we are free to pick.

- The change in $\bar{L}$ can be written as,

$$\triangle \bar{L} = \left\{ \frac{\partial L}{\partial \boldsymbol{x}} + \boldsymbol{\lambda}^T \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}} \right\} \triangle \boldsymbol{x} + \left\{ \frac{\partial L}{\partial \boldsymbol{u}} + \boldsymbol{\lambda}^T \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{u}} \right\} \triangle \boldsymbol{u} + \{\boldsymbol{c}(\boldsymbol{x}, \boldsymbol{u}) - \boldsymbol{\gamma}\}^T \triangle \boldsymbol{\lambda}$$

– Because we've introduced the Lagrange multipliers, we can treat all the parameters as if they were *independent*.

– So necessary conditions for a minimum are now:

$$\frac{\partial \bar{L}}{\partial \boldsymbol{x}} = \frac{\partial L}{\partial \boldsymbol{x}} + \boldsymbol{\lambda}^T \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}} = 0$$

$$\frac{\partial \bar{L}}{\partial \boldsymbol{u}} = \frac{\partial L}{\partial \boldsymbol{u}} + \boldsymbol{\lambda}^T \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{u}} = 0$$

$$\left\{\frac{\partial \bar{L}}{\partial \boldsymbol{\lambda}}\right\}^T = \boldsymbol{c}(\boldsymbol{x}, \boldsymbol{u}) - \boldsymbol{\gamma} = 0$$

which gives $(2n + m)$ equations in $(2n + m)$ unknowns.

- NOTE: Solving the first equation for $\boldsymbol{\lambda}$ and substituting this result into the second equation yields the same result as that derived using the Method of First Variations.
- Using Lagrange multipliers, we transform the problem from one of minimizing L subject to $\boldsymbol{c} = \boldsymbol{\gamma}$, to one of minimizing $\bar{L}$ *without constraints*.

Sufficient Conditions for a Local Minimum

- Using Lagrange multipliers, we can develop sufficient conditions for a minimum by expanding $\triangle \bar{L}$ to second order:

$$\triangle \bar{L} = \bar{L}_x \triangle \boldsymbol{x} + \bar{L}_u \triangle \boldsymbol{u} + \frac{1}{2} \begin{bmatrix} \triangle \boldsymbol{x}^T & \triangle \boldsymbol{u}^T \end{bmatrix} \begin{bmatrix} \bar{L}_{\boldsymbol{xx}} & \bar{L}_{\boldsymbol{xu}} \\ \bar{L}_{\boldsymbol{ux}} & \bar{L}_{\boldsymbol{uu}} \end{bmatrix} \begin{bmatrix} \triangle \boldsymbol{x} \\ \triangle \boldsymbol{u} \end{bmatrix}$$

where

$$\bar{L}_{\boldsymbol{xx}} = \frac{\partial^2 \bar{L}}{\partial \boldsymbol{x}^2} \qquad \bar{L}_{\boldsymbol{uu}} = \frac{\partial^2 \bar{L}}{\partial \boldsymbol{u}^2}$$

$$\bar{L}_{\boldsymbol{xu}} = \frac{\partial}{\partial \boldsymbol{u}} \left\{\frac{\partial L}{\partial \boldsymbol{x}}\right\}^T = \bar{L}_{\boldsymbol{ux}}$$

– For a stationary point,

$$\bar{L}_{\boldsymbol{x}} = \bar{L}_{\boldsymbol{u}} = 0$$

and

$$\triangle \boldsymbol{x} = -\boldsymbol{c}_{\boldsymbol{x}}^{-1}\boldsymbol{c}_{\boldsymbol{u}}\triangle \boldsymbol{u} + \text{h.o.t.}$$

– Substituting this expression into $\triangle \bar{L}$ and neglecting terms higher than second order yields:

$$\triangle \bar{L} = \frac{1}{2}\triangle \boldsymbol{u}^T \left[-\boldsymbol{c}_{\boldsymbol{u}}^T c_{\boldsymbol{x}}^{-T} \;\middle|\; I_{m\times m} \right] \begin{bmatrix} \bar{L}_{\boldsymbol{xx}} & \bar{L}_{\boldsymbol{xu}} \\ \bar{L}_{\boldsymbol{ux}} & \bar{L}_{\boldsymbol{uu}} \end{bmatrix} \begin{bmatrix} -c_{\boldsymbol{x}}^{-1}\boldsymbol{c}_{\boldsymbol{u}} \\ \hline I_{m\times m} \end{bmatrix} \triangle \boldsymbol{u}$$

$$= \frac{1}{2}\triangle \boldsymbol{u}^T \left[-\boldsymbol{c}_{\boldsymbol{u}}^T \boldsymbol{c}_{\boldsymbol{x}}^{-T} \;\middle|\; I_{m\times m} \right] \begin{bmatrix} -\bar{L}_{xx}c_x^{-1}\boldsymbol{c}_u + \bar{L}_{xu} \\ -\bar{L}_{ux}c_x^{-1}\boldsymbol{c}_u + \bar{L}_{uu} \end{bmatrix} \triangle \boldsymbol{u}$$

$$= \frac{1}{2}\triangle \boldsymbol{u}^T \left[\bar{L}_{\boldsymbol{uu}} - \bar{L}_{\boldsymbol{ux}} c_{\boldsymbol{x}}^{-1}\boldsymbol{c}_{\boldsymbol{u}} - \boldsymbol{c}_{\boldsymbol{u}}^T c_{\boldsymbol{x}}^{-T}\bar{L}_{\boldsymbol{xu}} + \boldsymbol{c}_{\boldsymbol{u}}^T c_{\boldsymbol{x}}^{-T}\bar{L}_{\boldsymbol{xx}} c_{\boldsymbol{x}}^{-1}\boldsymbol{c}_u \right] \triangle \boldsymbol{u}$$

– Thus we may write

$$\triangle \bar{L} = \frac{1}{2}\triangle \boldsymbol{u}^T \bar{L}_{uu}^* \triangle \boldsymbol{u}$$

– So, a sufficient condition for a local minimum is that

$\bar{L}_{\boldsymbol{uu}}^*$ is positive definite!

4.5: Equality Contraints: Examples

Example 4.2

- Returning to the previous problem of a rectangle within a circle, we have

$$\begin{aligned} \bar{L}_{xx} &= 2\lambda & \bar{L}_{yx} &= -4 & c_x &= 2x \\ \bar{L}_{xy} &= -4 & \bar{L}_{yy} &= 2\lambda & c_y &= 2y \end{aligned}$$

- Let y be the independent variable to write,

$$\bar{L}^*_{yy} = 2\lambda - \frac{-4y}{x} - \frac{-4y}{x} + \frac{2\lambda y^2}{x^2}$$

- At the minimum,

$$x = y = \frac{R}{\sqrt{2}}$$

$$\lambda = 2$$

$$\Rightarrow \bar{L}^*_{yy} = 16 > 0$$

Interpretation of Results

- Using first variations and/or Lagrange multipliers, we developed the following sufficient conditions for a local minimum:

$$L^*_u = L_u - L_x c_x^{-1} \boldsymbol{c}_u = 0 \qquad \boldsymbol{c} = \boldsymbol{\gamma}$$

$$\bar{L}^*_{uu} > 0$$

- We know that $L_{\boldsymbol{u}} = \dfrac{\partial L}{\partial \boldsymbol{u}}$, holding $\boldsymbol{x}$ constant, and $L_{\boldsymbol{x}} = \dfrac{\partial L}{\partial \boldsymbol{x}}$, holding $\boldsymbol{u}$ constant
  - Do we have a similar interpretation for $L^*_{\boldsymbol{u}}$?

$$\text{YES} \quad \Rightarrow L^*_{\boldsymbol{u}} = \frac{\partial L}{\partial \boldsymbol{u}}, \text{ holding } \boldsymbol{c} \text{ constant}$$

- So $L^*_{\boldsymbol{u}}$ automatically builds in the constraint relationship between $\boldsymbol{x}$ and $\boldsymbol{u}$
- Similarly,

$$L^*_{\boldsymbol{uu}} = \frac{\partial^2 L}{\partial \boldsymbol{u}^2}, \text{ holding } \boldsymbol{c} \text{ constant}$$

Additional Interpretations of Results:

- By introducing Lagrange multipliers, we found that we could solve equality constrained optimization problems using *unconstrained* optimization techniques
- It turns out that Lagrange multipliers serve another useful purpose as "cost sensitivity parameters"
  - Consider the adjoined cost developed previously,

$$\bar{L} = L + \boldsymbol{\lambda}^T(\boldsymbol{c} - \boldsymbol{\gamma})$$

$$\triangle \bar{L} = \left.\frac{\partial \bar{L}}{\partial \boldsymbol{x}}\right|_* \triangle \boldsymbol{x} + \left.\frac{\partial \bar{L}}{\partial \boldsymbol{u}}\right|_* \triangle \boldsymbol{u} - \boldsymbol{\lambda}^T \delta \boldsymbol{\gamma}$$

  - NOTE: the term $\delta\boldsymbol{\gamma}$ is included because we want to determine how the optimal solution changes when the constraints are changed by a *small* amount
  - But we know that at a minimum,

$$\left.\frac{\partial \bar{L}}{\partial \boldsymbol{x}}\right|_* = \left.\frac{\partial \bar{L}}{\partial \boldsymbol{u}}\right|_* = 0$$

  so any change in the cost, $\triangle \bar{L}$ (and hence $\triangle L$) is due to $\boldsymbol{\lambda}^T \delta \boldsymbol{\gamma} \;\Rightarrow$

$$\boldsymbol{\lambda}^T = -\frac{\partial L_{min}}{\partial \boldsymbol{\gamma}}$$

- Therefore, the Lagrange multipliers describe the *rate of change* of the optimal value of L with respect to the *constraints*

– Obviously, when the constraints change, the entire solution changes

– For small changes, we can perform a perturbation analysis to identify the new solution

– At the original minimum,

$$\bar{L}_{\boldsymbol{x}} = 0 \qquad \Rightarrow \quad \triangle \bar{L}_{\boldsymbol{x}} = 0$$

$$\bar{L}_{\boldsymbol{u}} = 0 \qquad \Rightarrow \quad \triangle \bar{L}_{\boldsymbol{u}} = 0$$

$$\boldsymbol{c} = \boldsymbol{\gamma} \qquad \Rightarrow \quad \triangle \boldsymbol{c} = \delta \boldsymbol{\gamma}$$

– Expanding these terms in a Taylor series expansion yields:

$$\triangle \bar{L}_{\boldsymbol{x}} = \bar{L}_{\boldsymbol{xx}} \triangle \boldsymbol{x} + \bar{L}_{\boldsymbol{xu}} \triangle \boldsymbol{u} + \boldsymbol{c}_{\boldsymbol{x}}^T \triangle \boldsymbol{\lambda} = 0$$

$$\triangle \bar{L}_{\boldsymbol{u}} = \bar{L}_{\boldsymbol{ux}} \triangle \boldsymbol{x} + \bar{L}_{\boldsymbol{uu}} \triangle \boldsymbol{u} + \boldsymbol{c}_{\boldsymbol{u}}^T \triangle \boldsymbol{\lambda} = 0$$

$$\boldsymbol{c}_x \triangle \boldsymbol{x} + \boldsymbol{c}_u \triangle \boldsymbol{u} = \delta \boldsymbol{\gamma}$$

$$\Rightarrow \; \triangle \boldsymbol{x} = \boldsymbol{c}_{\boldsymbol{x}}^{-1} \delta \boldsymbol{\gamma} - \boldsymbol{c}_{\boldsymbol{x}}^{-1} \boldsymbol{c}_u \triangle \boldsymbol{u} \qquad \triangle \boldsymbol{\lambda} = -\boldsymbol{c}_{\boldsymbol{x}}^{-T} \left\{ \bar{L}_{\boldsymbol{xx}} \triangle \boldsymbol{x} + \bar{L}_{\boldsymbol{xu}} \triangle \boldsymbol{u} \right\}$$

and substituting these results into the expression for $\triangle \bar{L}_{\boldsymbol{u}} = 0$ yields:

$$\triangle \boldsymbol{u} = -\Gamma \delta \boldsymbol{\gamma}$$

$$\Gamma = \bar{L}_{\boldsymbol{uu}}^{*-1} \left\{ \bar{L}_{\boldsymbol{ux}} - \boldsymbol{c}_{\boldsymbol{u}}^T c_{\boldsymbol{x}}^{-T} \bar{L}_{\boldsymbol{xx}} \right\} c_{\boldsymbol{x}}^{-1}$$

– NOTE: if $\bar{L}_{uu}^*$ exists, neighboring optimal solutions will exist.

Example 4.3

Continuing the previous example,

$$\begin{aligned} x &= y = \frac{R}{\sqrt{2}} & \bar{L}_{xx} &= \bar{L}_{yy} = 2\lambda & c_x &= 2x \\ \lambda &= 2 & \bar{L}_{xy} &= \bar{L}_{yx} = -4 & c_y &= 2y \end{aligned}$$

$$L = -4xy \quad \bar{L}_{yy} = 16$$

$$L^* = -2R^2$$

$$\triangle\gamma = \triangle R^2 \;\Rightarrow\; \text{change } R^2 \text{ by } \triangle R^2, \quad \triangle L^* = -2\triangle R^2 = -\lambda!$$

- What about $\triangle x$, $\triangle y$, $\triangle\lambda$?

$$\triangle x = \left\{\frac{1}{2x} - \frac{1}{4x}\right\} \triangle R^2 = \frac{1}{4x}\triangle R^2$$

$$\triangle y = -\frac{1}{16}\left\{-4 - \frac{2y}{2x}2\lambda\right\}\frac{\triangle R^2}{2x} = \frac{1}{4x}\triangle R^2$$

$$\triangle\lambda = -\frac{1}{2x}\left\{\frac{2\lambda}{4x} - \frac{4}{4x}\right\}\triangle R^2 = 0$$

- Does this agree with reality? To first order, YES!
  - λ is constant so $\triangle\lambda = 0$

$$x^* = y^* = \frac{R'}{\sqrt{2}} = \frac{1}{\sqrt{2}}\left(\sqrt{R^2 + \triangle R^2}\right)$$

$$= \frac{R}{\sqrt{2}}\left(\sqrt{1 + \frac{\triangle R^2}{R^2}}\right)$$

$$\approx \frac{R}{\sqrt{2}}\left\{1 + \frac{1}{2}\frac{\triangle R^2}{R^2}\right\}$$

where the last step makes use of the two-term Taylor series expansion of the square root.

$$\Rightarrow \triangle x = \triangle y = \frac{\triangle R^2}{2\sqrt{2}R} \quad \left\{= \frac{\triangle R^2}{4x}\right\}$$

Example 4.4

- Consider the two-parameter objective function

$$L(\boldsymbol{u}) = \frac{1}{2}u_1^2 + \frac{1}{2}u_2^2 - 2u_1 - 2u_2$$

with constraints:

$$c_1(\boldsymbol{u}) = u_1 + u_2 = 1$$
$$c_2(\boldsymbol{u}) = 2u_1 + 2u_2 = 6$$

- A graphic examination of the two constraint equations shows they are incompatible; i.e. the feasible region is empty.
- Consider a third constraint equation,

$$c_3(\boldsymbol{u}) = 6u_1 + 3u_2 = 6$$

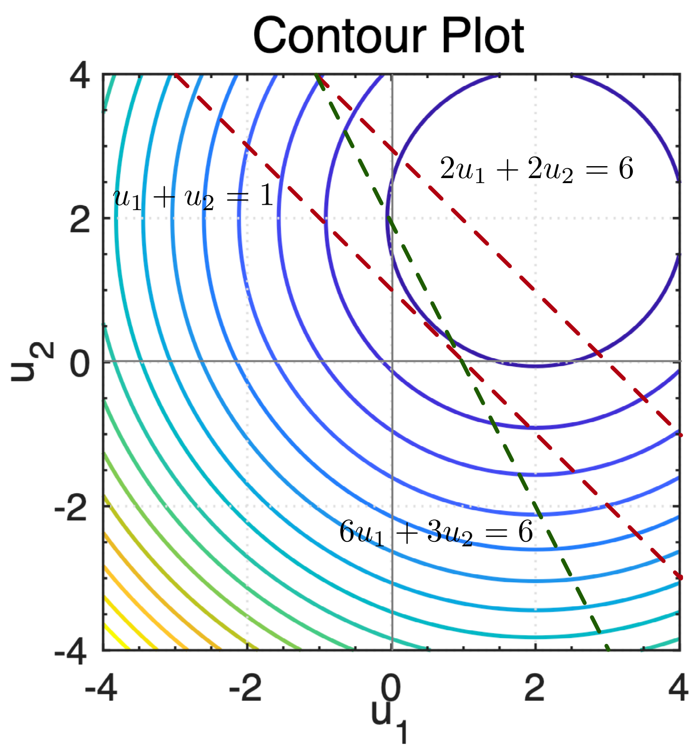

Figure 4.3: Example 4.4

  – The combination of constraints $c_1(\boldsymbol{u})$ and $c_3(\boldsymbol{u})$ produce a non-empty feasible region but show the general difficulty when the number of constrants n equals the number of paramters.

- The corresponding geometry is shown in Figure 4.3.

– Solving the constraint equations gives

$$\begin{bmatrix} 1 & 1 \\ 6 & 3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$$

$$\boldsymbol{u}^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- Let's now solve the general problem $L(\boldsymbol{u})$ with single constraint $c_1(\boldsymbol{u})$.
  - This time we have $n = 1$, giving a single dependent variable which we'll identify as u_2
- Forming the adjoined objective function we write

$$\bar{L} = \frac{1}{2}(u_1^2 + u_2^2) - 2(u_1 + u_2) + \lambda(u_1 + u_2 - 1)$$

- The necessary conditions for a minimum give:

$$\frac{\partial \bar{L}}{\partial u_1} = u_1 - 2 + \lambda = 0$$

$$\frac{\partial \bar{L}}{\partial u_2} = u_2 - 2 + \lambda = 0$$

$$\frac{\partial \bar{L}}{\partial \lambda} = u_1 + u_2 - 1 = 0$$

  - Solving this set of 3 equations in 3 unknowns we get

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$$

$$u_1^* = 0.5$$

$$u_2^* = 0.5$$
$$\lambda = 1.5$$

- Solving by the First Method of Variations, we use the equations,

$$\frac{\partial L}{\partial u_1} - \frac{\partial L}{\partial u_2}\left[\frac{\partial c}{\partial u_2}\right]^{-1}\frac{\partial c}{\partial u_1} = 0$$

$$c(u_1, u_2) = \gamma$$

to write

$$\begin{aligned}(u_1 - 2) - (u_2 - 2)\,(1)\,(1) &= 0\\ u_1 - 2 - u_2 + 2 &= 0\\ u_1 - u_2 &= 0\end{aligned}$$

and from the constraint equation,

$$u_1 + u_2 = 1$$

- Solving this set of 2 equations in 2 unknowns we get

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$u_1^* = 0.5$$
$$u_2^* = 0.5$$

as before.

- Checking for sufficiency, it is easy to show,

$$\bar{L}_{\boldsymbol{uu}}^* = 2 > 0$$

4.6: Numerical Algorithms

- As in the case of unconstrained optimization, the necessary and sufficient conditions for a local minimum may produce a set of equations that are too difficult to solve analytically.
  - The introduction of constraints complicates the problem much more quickly!
- So again, we must resort to numerical algorithms which, for the equality-constrained problem, will be developed by adapting "steepest-descent" procedures as presented previously

Second Order Gradient Method

- Unlike the unconstrained problem, we now have more to worry about than simply finding a minimizing parameter vector $\boldsymbol{u}^*$ (as if that wasn't enough!)
- In addition, we must also find $\boldsymbol{\lambda}$ and $\boldsymbol{x}$
- Fortunately, we have enough equations available to solve this problem $\rightarrow$
  - Given $\boldsymbol{u}$ $\Rightarrow$ Solve $\boldsymbol{c}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{\gamma}$ for $\boldsymbol{x}$

$$\Rightarrow \text{Solve } \frac{\partial \bar{L}}{\partial \boldsymbol{x}} = 0 \text{ for } \boldsymbol{\lambda}$$

$$\Rightarrow \text{Solve } \frac{\partial \bar{L}}{\partial \boldsymbol{u}} = 0 \text{ for new } \boldsymbol{u}$$

- Practical application of the theory:
  - As in the unconstrained problem, we'll select an initial $\boldsymbol{u}$ ($\boldsymbol{u}^{(k)}$) that we'll *assume* to be correct.

- Then, solving for $\triangle \boldsymbol{x}$, we can compute an update to $\boldsymbol{x}^{(k)}$,

$$\boldsymbol{c}(\boldsymbol{x},\, \boldsymbol{u}) \approx \boldsymbol{c}(\boldsymbol{x}^{(k)},\, \boldsymbol{u}^{(k)}) + \left.\frac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}}\right|_k \triangle \boldsymbol{x} = \boldsymbol{\gamma}$$

$$\Rightarrow \triangle \boldsymbol{x} = -\left\{ \left.\frac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}}\right|_k \right\}^{-1} \left[\boldsymbol{c}(\boldsymbol{x}^{(k)},\, \boldsymbol{u}^{(k)}) - \boldsymbol{\gamma} \right]$$

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \triangle \boldsymbol{x}$$

- By iterating through this process, we find the value of $\boldsymbol{x}$ which satisfies the constraints for the given $\boldsymbol{u}$.
- Notice, however, that we must select an initial $\boldsymbol{x}$ to start the iteration process.
- Now, solving for $\boldsymbol{\lambda}$ $\Rightarrow$

$$\frac{\partial \bar{L}}{\partial \boldsymbol{x}} = \frac{\partial L}{\partial \boldsymbol{x}} + \boldsymbol{\lambda}^T \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}} = \boldsymbol{0}^T$$

$$\Rightarrow \boldsymbol{\lambda}^T = -\frac{\partial L}{\partial \boldsymbol{x}} \left\{ \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}} \right\}^{-1}$$

- If at this point $\boldsymbol{u}$ and $\boldsymbol{x}$ are truly the optimal solution, then $\partial \bar{L}/\partial \boldsymbol{u} = \boldsymbol{0}^T$
- But, if $\partial \bar{L}/\partial \boldsymbol{u} \neq \boldsymbol{0}^T$, then we can use the following procedure to iterate for $\boldsymbol{u}$,

$$\left.\frac{\partial \bar{L}}{\partial \boldsymbol{u}}^T\right|_{c=\boldsymbol{\gamma},\, \boldsymbol{u}+\triangle \boldsymbol{u}} = \left.\frac{\partial \bar{L}}{\partial \boldsymbol{u}}^T\right|_{c=\boldsymbol{\gamma},\, \boldsymbol{u}} + \bar{L}^*_{uu} \triangle \boldsymbol{u} = \boldsymbol{0}$$

where

$$\bar{L}^*_{\boldsymbol{uu}} = \bar{L}_{\boldsymbol{uu}} - L_{\boldsymbol{ux}} \boldsymbol{c}_{\boldsymbol{x}}^{-1} \boldsymbol{c}_u - \boldsymbol{c}_{\boldsymbol{u}}^T \boldsymbol{c}_{\boldsymbol{x}}^{-T} \bar{L}_{\boldsymbol{xu}} + \boldsymbol{c}_{\boldsymbol{u}}^T \boldsymbol{c}_{\boldsymbol{x}}^{-T} \bar{L}_{\boldsymbol{xx}} \boldsymbol{c}_{\boldsymbol{x}}^{-1} \boldsymbol{c}_{\boldsymbol{u}}$$

$$\Rightarrow \triangle \boldsymbol{u} = -\bar{L}^{*-1}_{\boldsymbol{uu}} \frac{\partial \bar{L}}{\partial \boldsymbol{u}}^T$$

– It's important to exercise care about how you calculate $\bar{L}_{\boldsymbol{uu}}$, $\bar{L}_{\boldsymbol{ux}}$, $\bar{L}_{\boldsymbol{xu}}$, and $\bar{L}_{\boldsymbol{xx}}$:

$$\frac{\partial \bar{L}}{\partial \boldsymbol{x}} = \frac{\partial L}{\partial \boldsymbol{x}} + \boldsymbol{\lambda}^T \frac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}}$$

$$\Rightarrow \quad \frac{\partial^2 \bar{L}}{\partial \boldsymbol{x}^2} = \frac{\partial^2 L}{\partial \boldsymbol{x}^2} + \sum_{i=1}^{n} \lambda_i \frac{\partial^2 c_i}{\partial \boldsymbol{x}^2}$$

– Note that our matrix notation has broken down to some extent with respect to the term

$$\sum_{i=1}^{n} \lambda_i \frac{\partial^2 c_i}{\partial \boldsymbol{x}^2}$$

where the terms $\dfrac{\partial c_i}{\partial \boldsymbol{x}^2}$ are matrix-valued (second derivative of a scalar by a vector – think Hessian) and the summation comprises the linear combination of these matrix elements scaled by the elements of the Lagrange multiplier vector.

First-Order Gradient Method

- The same arguments used to justify the first-order algorithm for the unconstrained problem also apply here
- So, $\boldsymbol{x}$ and $\boldsymbol{\lambda}$ can be identified using the iterative procedure introduced above; but $\boldsymbol{u}$ is updated as follows:

$$\triangle \boldsymbol{u} = -K \frac{\partial \bar{L}}{\partial \boldsymbol{u}}^T$$

where K is a positive scalar when searching for a minimum.

Prototype Algorithm

1. Guess $\boldsymbol{x}$ and $\boldsymbol{u}$

2. Compute $\boldsymbol{c}-\boldsymbol{\gamma}$; If $\|\boldsymbol{c}-\boldsymbol{\gamma}\|<\epsilon_1$, go to (5.)
3. Compute $\dfrac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}}$
4. Update $\boldsymbol{x}=\boldsymbol{x}^{(k)}-\left\{\dfrac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}}\right\}^{-1}[\boldsymbol{c}-\boldsymbol{\gamma}]$; go to (2.)
5. Compute L
6. Compute $\dfrac{\partial L}{\partial \boldsymbol{x}}, \dfrac{\partial L}{\partial \boldsymbol{u}}, \dfrac{\partial \boldsymbol{c}}{\partial \boldsymbol{u}}$
7. Compute $\boldsymbol{\lambda}^T=-\dfrac{\partial L}{\partial \boldsymbol{x}}\left\{\dfrac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}}\right\}^{-1}$
8. Compute $\dfrac{\partial \bar{L}}{\partial \boldsymbol{u}}=\dfrac{\partial L}{\partial \boldsymbol{u}}+\boldsymbol{\lambda}^T\dfrac{\partial \boldsymbol{c}}{\partial \boldsymbol{u}}$; If $\left\|\dfrac{\partial \bar{L}}{\partial \boldsymbol{u}}\right\|<\epsilon_2$, stop!
9. Update $\boldsymbol{u}^{(k+1)}=\boldsymbol{u}^{(k)}-K\dfrac{\partial \bar{L}}{\partial \boldsymbol{u}}$ or $\boldsymbol{u}^{(k+1)}=\boldsymbol{u}^{(k)}-\bar{L}_{\boldsymbol{uu}}^{*-1}\dfrac{\partial \bar{L}}{\partial \boldsymbol{u}}^T$; go to (2.)

4.7: Inequality-Constraints: Scalar Parameter

- Our goal in the equality constrained parameter optimization problem discussed in the last section was to find a set of parameters

$$\mathbf{y}^T = \begin{bmatrix} y_1 & y_2 & \cdots & y_p \end{bmatrix}$$

which minimizes $L(\mathbf{y})$ subject to a set of constraints:

$$c_i(\mathbf{y}) = \gamma_i, \quad i = 1, 2, \ldots, n$$

where $n < p$

  - Note, if $n = p$, the problem is completely specified and no optimization is necessary.

- Now, we want to examine the problem of selecting $\mathbf{y}$ to minimize $L(\mathbf{y})$ subject to a set of *inequality* constraints:

$$c_i(\mathbf{y}) \le \gamma_i, \quad i = 1, 2, \ldots, n$$

where n is no longer related to p (i.e., $n < p$ or $n \ge p$)

  - Why is n not related to p? Because some constraints may be *inactive* $\Rightarrow$ meaning no constraint exists at all.

- To solve this problem, we will not split the parameters into state (constraint) and decision variables (as done previously) for the simple reason that it can't be done for the case where the number of constraints is greater than the number of parameters.

- We will approach this problem by focusing next on the simplest case and then generalizing to more complicated situations as was done for the equality-constrained problem.

Scalar Parameter / Scalar Constraint

GOAL: Minimize $L(y)$ subject to the constraint $c(y) \le 0$

Case 1: Constraint inactive (i.e., $c(y^*) < 0$)

- Constraint can be ignored
- So the problem is identical to the unconstrained optimization problem discussed in the previous section.

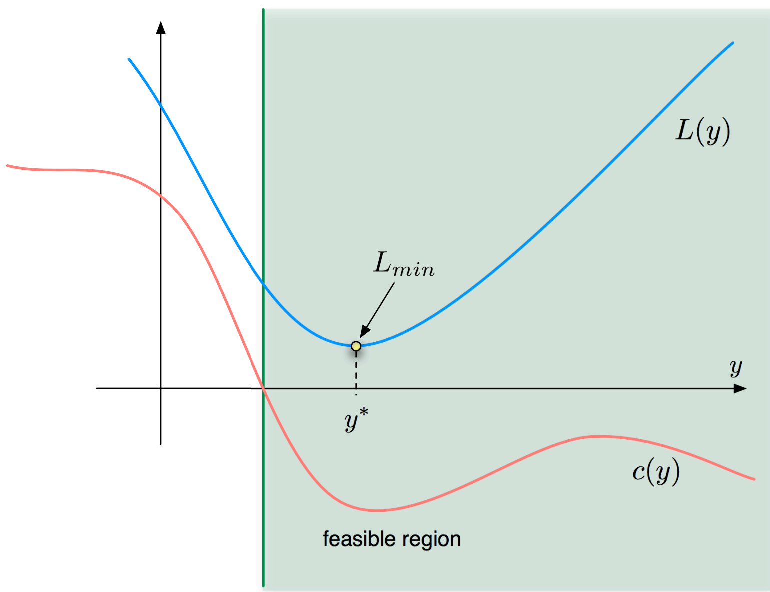

Figure 4.4: Constrained Optimization: Scalar Function Case 1

Case 2: Constraint active (i.e., $c(y^*) = 0$)

- If y^* is the optimal solution, then for all admissible perturbations away from y^* the following conditions must exist:

$$\triangle L \approx {}^{dL}/_{dy}\big|_{y^*} \triangle y \ge 0 \quad \& \quad \triangle c \approx {}^{dc}/_{dy}\big|_{y^*} \triangle y \le 0$$

$$\Rightarrow \quad \operatorname{sgn} {}^{dL}/_{dy} = -\operatorname{sgn} {}^{dc}/_{dy} \qquad \text{OR} \qquad {}^{dL}/_{dy} = 0$$

- These two conditions can be neatly summarized into a single relationship using a Lagrange multiplier:

$$ {}^{dL}/_{dy} + \lambda {}^{dc}/_{dy} = 0, \quad \lambda \ge 0$$

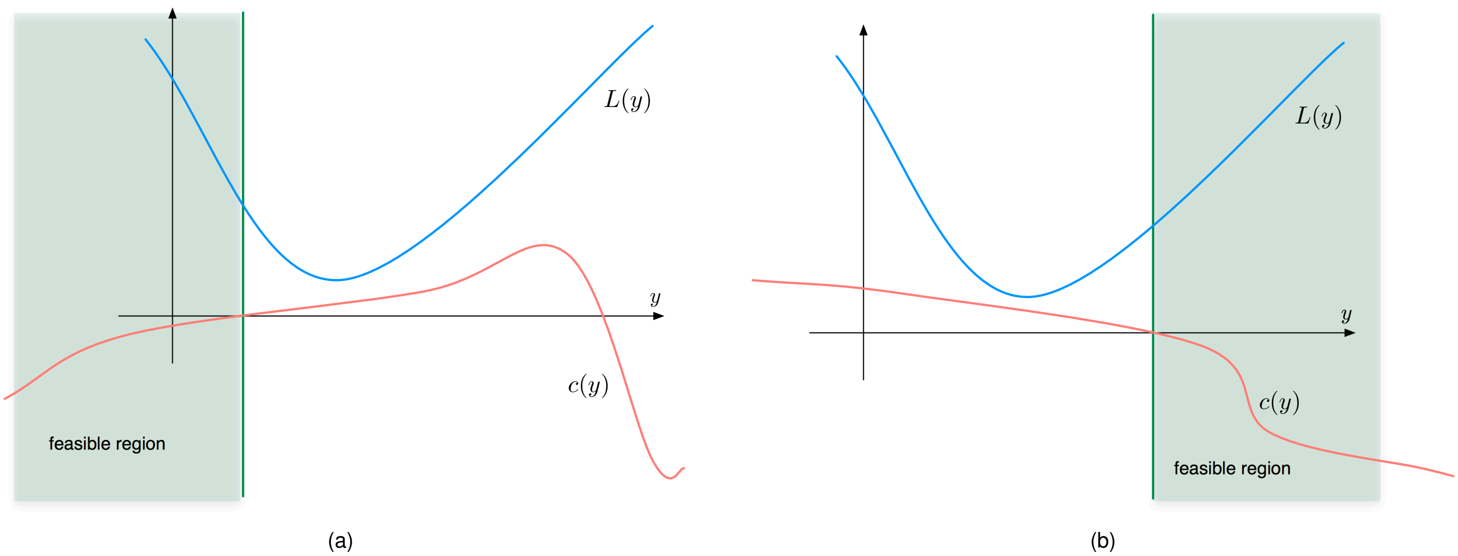

Figure 4.5: Constrained Optimization: Scalar Function Case 2

$$\left.{}^{dc}\!/_{dy}\right|_{y^*} > 0 \;\Rightarrow\; \triangle y_{feas} < 0 \qquad\qquad \left.{}^{dc}\!/_{dy}\right|_{y^*} < 0 \;\Rightarrow\; \triangle y_{feas} > 0$$

$$\triangle y_{feas} < 0 \;\Rightarrow\; \left.{}^{dL}\!/_{dy}\right|_{y^*} \leq 0 \qquad\qquad \triangle y_{feas} > 0 \;\Rightarrow\; \left.{}^{dL}\!/_{dy}\right|_{y^*} \geq 0$$

- In each of these cases, y^* occurs at the constraint boundary, and
$${}^{dL}\!/_{dy} + \lambda\, {}^{dc}\!/_{dy} = 0 \quad \text{for } \lambda > 0!$$

- The conditions for a minimum in both Case 1 and Case 2 can, in fact, be handled analytically using the same cost function:
$$\bar{L} = L + \lambda c$$
  - The necessary conditions for a minimum become:
  $${}^{d\bar{L}}\!/_{dy} = 0 \qquad \text{and} \qquad c(y) \leq 0$$
  where $\lambda \geq 0$ for $c(y) = 0$ and $\lambda = 0$ for $c(y) < 0$
  - We call the the relationship,
  $$\lambda \cdot c\,(y) = 0$$
  the *complimentarity condition*.

Example

- Consider the function $L(u) = \dfrac{1}{2}u^2$
- Unconstrained minimum:

$$ {}^{dL}\!/_{du} = 0 \quad \Rightarrow \quad u = 0 $$

1. $u \le k,\ k > 0$

$$ \frac{d\bar{L}}{du} = 0 $$

$$ \Rightarrow u + \lambda = 0 $$

$$ \lambda = 0 \quad \Rightarrow\ u = 0 \quad \text{satisfies constraint} $$

2. $u \le k,\ k < 0$

$$ \frac{d\bar{L}}{Lu} = 0 $$

$$ \lambda = 0 \quad \Rightarrow\ u = 0 \quad \text{no good} $$

$$ \lambda \ne 0 \quad \Rightarrow\ u = k\ \Rightarrow\ \lambda = -k > 0! $$

4.8: Inequality Constraints: Vector Parameter

Vector Parameter / Scalar Constraint

GOAL: Minimize $L(\boldsymbol{y})$ subject to $c(\boldsymbol{y}) \leq 0$

- For this problem, the results developed above must simply be reinterpreted:
  - For all admissable $\triangle \boldsymbol{y}$, $\triangle L \approx {}^{\partial L}/_{\partial \boldsymbol{y}} \triangle \boldsymbol{y} \geq 0$
  - If $\boldsymbol{y}^*$ exists on a constraint boundary, then

$$\triangle c \approx {}^{\partial c}/_{\partial \mathrm{y}} \triangle \boldsymbol{y} \leq 0 \quad \text{for all admissible } \triangle \boldsymbol{y}$$

$$\Rightarrow \frac{\partial L}{\partial \boldsymbol{y}} + \lambda \frac{\partial c}{\partial \boldsymbol{y}} = 0 \quad (\lambda \geq 0)$$

  - Thus, either ${}^{\partial L}/_{\partial \boldsymbol{y}} = 0$ or ${}^{\partial L}/_{\partial \boldsymbol{y}}$ is parallel to ${}^{\partial c}/_{\partial \boldsymbol{y}}$ and in the opposite direction at $\boldsymbol{y}^*$.

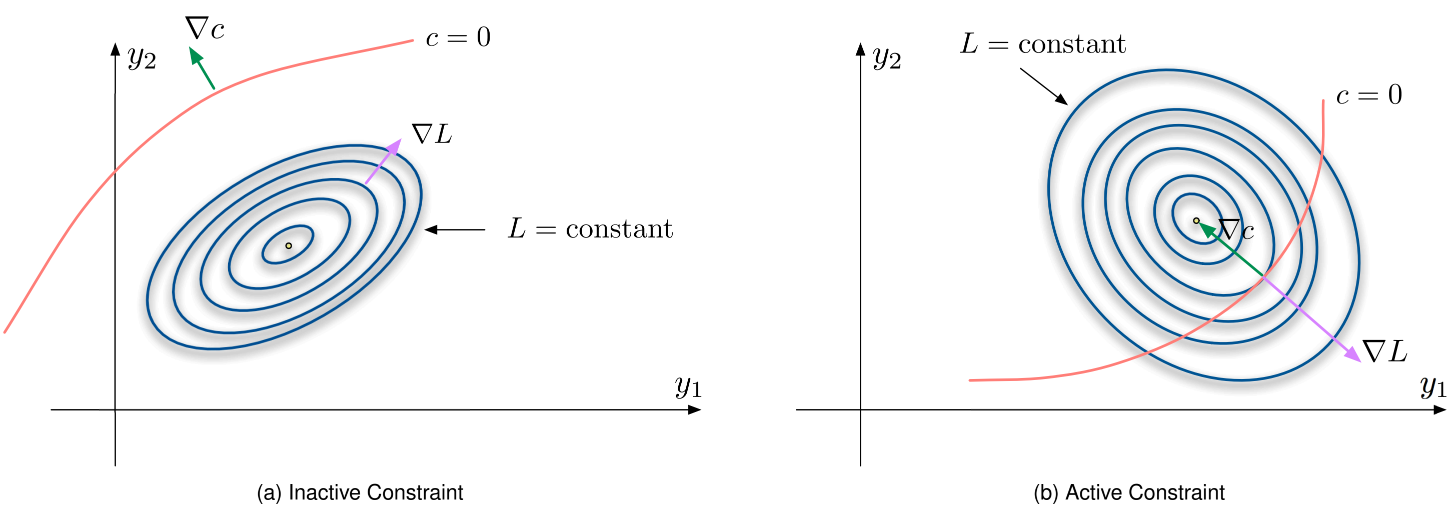

Figure 4.6

- As for the scalar/scalar case, we can handle this problem analytically using the following cost function:

$$\bar{L} = L + \lambda c$$

- The necessary conditions for a minimum are:

$$\frac{\partial \bar{L}}{\partial \boldsymbol{y}} = 0 \quad \text{and} \quad c(\boldsymbol{y}) \le 0$$

where $\lambda \ge 0$ for $c(\boldsymbol{y}) = 0$ and $\lambda = 0$ for $c(\boldsymbol{y}) < 0$

Example: Package Constraint Problem

- Maximize the volume of a rectangular box under the inequality dimension constraint: $2(x + y) + z \le D$

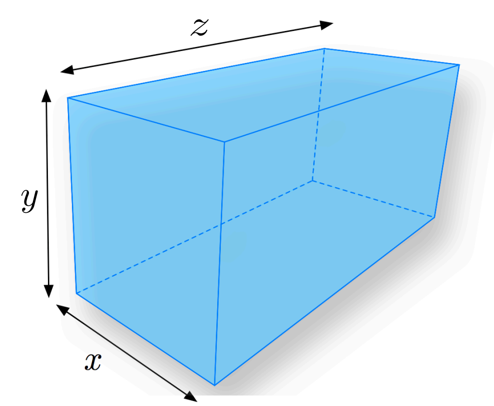

Figure 4.7: Example: Package Constraint Problem

- Objective function:

$$L = -xyz$$

- Solving,

$$\bar{L} = -xyz + \lambda(2x + 2y + z - D)$$

$$\frac{\partial L}{\partial \boldsymbol{u}} = \left[\, -yz + 2\lambda \;\middle|\; -xz + 2\lambda \;\middle|\; -xy + \lambda \,\right]$$

- $\lambda = 0 \;\Rightarrow$ either $x = 0$ or $y = 0 \Rightarrow$ volume is not maximized
- $\lambda \neq 0 \;\Rightarrow$

1. $yz = 2\lambda$
2. $xz = 2\lambda$
3. $xy = \lambda$
4. $2x + 2y + z = D$

- Combining (1) and (2) gives $x = y$
- Combining (3) with this result gives $\lambda = x^2$
- From (2) we get $z = 2x$
- From (4) we compute $x = D/6$

$$\Rightarrow \quad x^* = y^* = D/6$$
$$z^* = D/3$$

Vector Parameter / Vector Constraint

- The problem is still exactly the same as before, but the complexity of the solution continues to increase.
  - If $\boldsymbol{y}^*$ is the optimum, then
  $$\triangle L \approx \partial L/\partial \boldsymbol{y}\big|_{\boldsymbol{y}^*} \triangle \boldsymbol{y} \geq 0$$
  for all admissible $\triangle \boldsymbol{y}$
  - And if y^* lies on a constraint boundary, then
  $$\triangle c_i \approx \partial c_i/\partial \boldsymbol{y}\big|_{\boldsymbol{y}^*} \triangle \boldsymbol{y} \leq 0 \quad \{i = 1, 2, \ldots, q\}$$
  - Again, these conditions can be summarized using Lagrange multipliers (only now we need q of them):
  $$\partial L/\partial \boldsymbol{y} + \sum_{i=1}^{q} \lambda_i \left(\partial c_i/\partial \boldsymbol{y}\right) = 0 \quad \lambda_i > 0$$
  or in vector notation,
  $$\partial L/\partial \boldsymbol{y} + \boldsymbol{\lambda}^T \left(\partial \boldsymbol{c}/\partial \boldsymbol{y}\right) = 0 \quad \boldsymbol{\lambda} \geq \boldsymbol{0}$$

- The last equation above is known as a *Kuhn-Tucker* condition; we'll hear more about this later.
- Interpretation:
  1. If no constraints are active, $\boldsymbol{\lambda} = 0$ and $\partial L / \partial \boldsymbol{y} = 0$
  2. If some constraints are active, $\partial L / \partial \boldsymbol{y}$ must be a negative linear combination of the appropriate gradients ($\partial c_i / \partial \boldsymbol{y}$)
     (a) Physically, this means that $-\partial L / \partial \boldsymbol{y}$ must lie inside a cone formed by the active constraint gradients (i.e., $\partial L / \partial \boldsymbol{y}$ at a minimum must be pointed so that any decrease in L can only be achieved by violating the constraints).

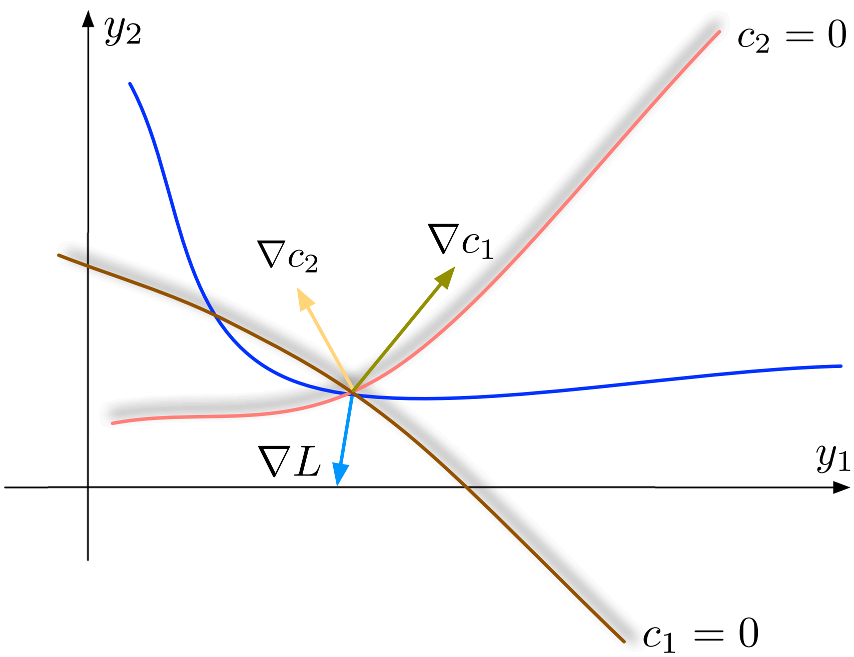

Figure 4.8: Vector Parameter / Vector Constraint

- Summary:
  - Define: $\bar{L} = L + \boldsymbol{\lambda}^T \boldsymbol{c}$
  - Necessary conditions for a minimum are:

$$\partial L/\partial \boldsymbol{y} = 0 \qquad c_i(\boldsymbol{y}) \le 0 \quad \{i = 1, 2, \ldots, q\}$$

where $\lambda_i \ge 0$ for $c_i(\boldsymbol{y}) = 0$ and $\lambda_i = 0$ for $c_i(\boldsymbol{y}) < 0$

– Question: How many constraints can be active?

4.9: Linear Programming

- The simplest type of constrained optimization problem occurs when the objective function and the constraint functions are all *linear* functions of $\boldsymbol{u}$: these are known as *Linear Programming* problems
- Standard linear programming problem may be stated as:

$$\min_{\boldsymbol{x}} \; f(\boldsymbol{x}) = \boldsymbol{c}^T \boldsymbol{x}$$

$$\text{subject to } A\boldsymbol{x} = \boldsymbol{b}, \; \boldsymbol{x} \geq \boldsymbol{0}$$

- Here, matrix A is dimension $(m \times n)$ and $m \leq n$ (usually)
- Coefficients $\boldsymbol{c}$ are often referred to as *costs*
- Although the standard problem is an equality constrained problem, inequality constraints may be accommodated by introducing additional variables.
  - For example,

$$x_l = \boldsymbol{a}_i^T \boldsymbol{x} - b_i$$

where

$$x_l \geq 0$$

- One important feature of linear programming minimization problems is that they *require* constraints, since linear objective functions have no minima!

Example (problem set-up):

$$\begin{aligned} \text{minimize} \quad & x_1 + 2x_2 + 3x_3 + 4x_4 & \\ \text{subject to} \quad & x_1 + x_2 + x_3 + x_4 & = 1 \\ & x_1 + x_3 - 3x_4 & = 1/2 \end{aligned}$$

$$x1 \geq 0,\ x_2 \geq 0\ x_3 \geq 0,\ x_4 \geq 0$$

- For this example, the vectors set up as

$$\boldsymbol{c}^T = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$$
$$\boldsymbol{a}_1^T = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$$
$$\boldsymbol{a}_2^T = \begin{bmatrix} 1 & 0 & 1 & -3 \end{bmatrix}$$

  and

$$\boldsymbol{b} = \begin{bmatrix} 1 \\ 1/2 \end{bmatrix}$$

- In this case, equations $A\boldsymbol{x} = \boldsymbol{b}$ determine a unique solution, so the solution is completely determined by the constraints.
- More commonly, $m < n$
  - In this case $A\boldsymbol{x} = b$ is underdetermined leaving $n - m$ degrees of freedom

Example

- For the example above, we can rearrange the constraint equation to give:

$$x_1 = \frac{1}{2} - x_3 + 3x_4$$
$$x_2 = \frac{1}{2} - 4x_4$$

- Or, alternatively we can write:

$$x_1 = \frac{7}{8} - \frac{3}{4}x_2 - x_3$$

$$x_4 = \frac{1}{8} - \frac{1}{4}x_2$$

- The objective function $\boldsymbol{c}^T\boldsymbol{x}$ is linear, so it does not contains the curvature needed to give rise to a minimum point
  - A minimum point must be created by the conditions $x_i \geq 0$ becoming active on the boundary of the feasible region.
- Substituting the second form of these equations into the main problem statement allows us to write

$$f(\boldsymbol{x}) = x_1 + 2x_2 + 3x_3 + 4x_4 = \frac{11}{8} + \frac{1}{4}x_2 + 2x_3$$

- Obviously this function has no minimum unless we impose the bounds $x_2 \geq 0$ and $x_3 > 0$; in this case $x_2 = x_3 = 0$ and the minimum is $f_{\min} = {}^{11}\!/_{8}$.

Example

- Consider the simple set of conditions:

$$x_1 + 2x_2 = 1, \quad x_1 \geq 0,\ x_2 \geq 0$$

- The first of these expressions gives the equation of a line:

$$x_2 = -{}^{1}\!/_{2}x_1 + {}^{1}\!/_{2}$$

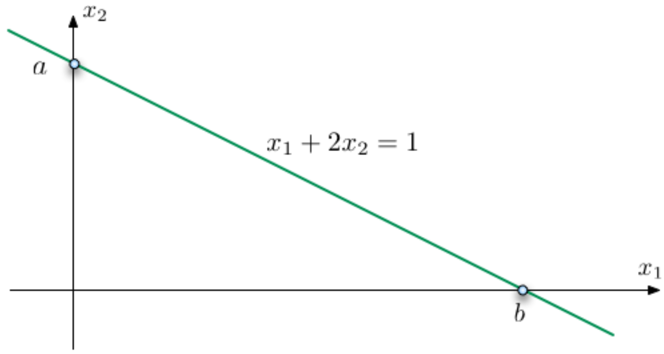

Figure 4.9: Linear Programming Example

- The feasible region is the line joining points $a = [0,\ 1/2]$ and $b = [1,\ 0]$
- Whenever the objective function is linear, the solution must occur at either a or b with $x_1 = 0$ or $x_2 = 0$
  - In the case where $f(\boldsymbol{x}) = x_1 + 2x_2$, then any point on the line segment is a solution (*non-unique* solution)
- Summarizing:
  - A solution of a linear programming problem always exists at one particular *extreme point* or *vertex* of the feasible region
  - At least $n - m$ variables have value *zero.*
  - The remaining m variables are determined uniquely from the equations $A\boldsymbol{x} = b$ and $\boldsymbol{x} \geq 0$.

4.10 Linear Programming: Simplex Algorithm

- The main challenge in solving linear programming problems is finding which $n - m$ variables equal zero at the solution.
- One popular method of solution is the SIMPLEX METHOD, which tries different sets of possibilites in a systematic manner.
  - The method generates a sequence of feasible points $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots$ which terminates at a solution.
  - Each iterate $\boldsymbol{x}^{(k)}$ is an extreme point.
- We define:
  - The set of *nonbasic variables* (set $N^{(k)}$) as the $n - m$ variables having zero value at $\boldsymbol{x}^{(k)}$.
  - The set of *basic variables* (set $B^{(k)}$) as the remaining m variables having non-zero value.
- Parameter vector $\boldsymbol{x}$ is partitioned so that the basic variables are the first m elements:

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_N \end{bmatrix}$$

- Likewise, we correspondingly partition A :

$$A = \left[\begin{array}{c|c} A_B & A_N \end{array}\right]$$

- Then we can write the constraint equations as:

$$\left[\begin{array}{c|c} A_B & A_N \end{array}\right] \begin{bmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_N \end{bmatrix} = A_B \boldsymbol{x}_B + A_N \boldsymbol{x}_N = \boldsymbol{b}$$

- Also, since $\boldsymbol{x}_N^{(k)} = 0$, we can write

$$\boldsymbol{x}^{(k)} = \begin{bmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_N \end{bmatrix}^{(k)} = \begin{bmatrix} \hat{\boldsymbol{b}} \\ \boldsymbol{0} \end{bmatrix}$$

where

$$\hat{\boldsymbol{b}} = A_B^{-1}\boldsymbol{b} \quad \text{and } \hat{\boldsymbol{b}} \geq \boldsymbol{0}$$

Example

- Returning to our previous example, let us choose $B = \{1, 2\}$ and $N = \{3, 4\}$ (i.e., variables x_1 and x_2 are basic, and x_3 and x_4 are nonbasic)
- We then have,

$$A_B = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \qquad A_N = \begin{bmatrix} 1 & 1 \\ 1 & -3 \end{bmatrix},$$

and

$$\hat{\boldsymbol{b}} = A_B^{-1}\boldsymbol{b} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \geq \boldsymbol{0}$$

- Since A_B is nonsingular and $\hat{\boldsymbol{b}} \geq \boldsymbol{0}$, this choice of B and N gives a basic feasible solution
- Solving for the value of the objective function,

$$\hat{f} = \boldsymbol{c}^T \boldsymbol{x}^{(k)} = \boldsymbol{c}_B^T \hat{\boldsymbol{b}}$$

  where we have partitioned $\boldsymbol{c}$ as

$$\boldsymbol{c} = \begin{bmatrix} \boldsymbol{c}_B \\ \boldsymbol{c}_N \end{bmatrix}$$

- Since

$$\mathbf{c}_B = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

we compute

$$\hat{f} = 1.5$$

- Now examine whether or not our basic feasible solution is optimal.
  - The essential idea here is to eliminate the basic variables from the objective function and reduce it to a function of nonbasic variables only.
  - This way we can determine whether an increase in any nonbasic variable will reduce the objective function further.
- Reducing $f(\mathbf{x})$ we can write

$$f(x) = \mathbf{c}_B^T \mathbf{x}_B + \mathbf{c}_N^T \mathbf{x}_N$$

- Since

$$\begin{aligned} A_B \mathbf{x}_B + A_N \mathbf{x}_N &= \mathbf{b} \\ A_B \mathbf{x}_B &= \mathbf{b} - A_N \mathbf{x}_N \\ \mathbf{x}_B &= A_B^{-1} \mathbf{b} - A_B^{-1} A_N \mathbf{x}_N \end{aligned}$$

- Substituting $\mathbf{x}_B$ above, we can write,

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{c}_B^T \hat{\mathbf{b}} - \mathbf{c}_B^T A_B^{-1} A_N \mathbf{x}_N + \mathbf{c}_N^T \mathbf{x}_N \\ &= \mathbf{c}_B^T \hat{\mathbf{b}} + \left[\mathbf{c}_N^T - \mathbf{c}_B^T A_B^{-1} A_N \right] \mathbf{x}_N \\ &= \hat{f} + \hat{\mathbf{c}}_N^T \mathbf{x}_N \end{aligned}$$

where (upon transposing) the *reduced cos*t can be written,

$$\hat{\mathbf{c}}_N = \mathbf{c}_N - A_N^T A_B^{-T} \mathbf{c}_B$$

- Now, with $f(\boldsymbol{x})$ in terms of $\boldsymbol{x}_N$ it is straightforward to check the conditions for which $f(\boldsymbol{x}_N)$ can be reduced.
  - Note here that although $\boldsymbol{x}_N = \mathbf{0}$ at a basic feasible solution, in general, $\boldsymbol{x}_N \geq \mathbf{0}$
  - So we define the optimality test:
$$\hat{\boldsymbol{c}}_N \geq \mathbf{0}$$
  - If the optimality test is satisfied, our solution is optimal and we terminate the algorithm (since $\boldsymbol{x}_N \geq \mathbf{0}$, $f(\boldsymbol{x})$ must increase).
  - If it is not satisfied, we have more work to do.
- Denote $\hat{\boldsymbol{c}}_N$ by
$$\hat{\boldsymbol{c}}_N = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \hat{c}_q \\ \vdots \end{bmatrix}$$
- Choose variables x_q for which $\hat{c}_q < 0$, which implies $f(\boldsymbol{x}_N)$ is decreased by increasing x_q (usually choose most negative $\hat{c}_q$)
  - As x_q increases, in order to keep $A\boldsymbol{x} = \boldsymbol{b}$ satisfied, $\boldsymbol{x}_B$ changes according to
$$\boldsymbol{x}_B = A_B^{-1}(\boldsymbol{b} - A_N\boldsymbol{x}_N) = \hat{\boldsymbol{b}} - A_B^{-1}A_N\boldsymbol{x}_N$$
  - In general, since x_q is the only nonbasic variable changing,
$$\begin{aligned} \boldsymbol{x}_B &= \hat{\boldsymbol{b}} - A_B^{-1}\boldsymbol{a}_q x_q, \\ &= \hat{\boldsymbol{b}} - \boldsymbol{d}\,x_q \end{aligned}$$
    where $\boldsymbol{a}_q$ is the column of matrix A corresponding to q.

- In this development, $\boldsymbol{d}$ behaves like a derivative of $\boldsymbol{x}_B$ w.r.t. x_q
- So, our approach is to increase x_q (thereby decreasing $f(\boldsymbol{x})$) until another element of $\boldsymbol{x}_B$ reaches 0.
  - It is clear that x_i becomes 0 when
$$x_q = \frac{\hat{b}_i}{-d_i}$$
  - Since the amount by which x_q can be increased is limited by the first basic variable to become 0 , we can state the *ratio test*:
$$\frac{\hat{b}_p}{-d_p} = \min_{i \in B} \frac{\hat{b}_i}{-d_i}$$
    where $d_i < 0$.
  - Geometrically, the increase in x_q and the corresponding change to $\boldsymbol{x}_B$ causes a move along an *edge* of the feasible region.
  - When a new element x_i reaches 0, the new sets $N^{(k+1)}$ and $B^{(k+1)}$ are re-ordered and an iteration of the simplex method is complete.

Example

- Finshing our example, we can express $f(\boldsymbol{x})$ in terms of the reduced cost and $\boldsymbol{x}_N$ as follows:
$$\begin{aligned} f(\boldsymbol{x}) &= \hat{f} + \hat{\boldsymbol{c}}_N^T \boldsymbol{x}_N \\ &= 1.5 + \begin{bmatrix} 2 & -1 \end{bmatrix} \boldsymbol{x}_N \end{aligned}$$
- Since $\hat{\boldsymbol{c}}_N$ does not satisfy the optimality test, our basic feasible solution is not an optimal one
- The negative value of $\hat{\boldsymbol{c}}_N$ corresponds to $q = 4$, i.e., $\hat{c}_4 = -1$

- So, with

$$\boldsymbol{a}_4 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

we can compute

$$\boldsymbol{d}_4 = -A_B^{-1}\boldsymbol{a}_4 = -\begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$$

- Therefore,

$$\boldsymbol{x}_B = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + \begin{bmatrix} 3 \\ -4 \end{bmatrix} x_4$$

from which we find that the max value $x_4 = 1/8$ brings the second element of $\boldsymbol{x}_B$ to zero.

- Now, we re-partition:

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_N \end{bmatrix} = \begin{bmatrix} x_1 \\ x_4 \\ \hline x_2 \\ x_3 \end{bmatrix}$$

$$A = \begin{bmatrix} A_B & A_N \end{bmatrix} = \left[\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & -3 & 0 & 1 \end{array}\right]$$

from which we compute

$$\hat{\boldsymbol{b}} = \begin{bmatrix} 0.8750 \\ 0.1250 \end{bmatrix} \geq \boldsymbol{0}$$

and

$$\hat{\boldsymbol{c}}_N = \begin{bmatrix} 0.25 \\ 2.0 \end{bmatrix} > 0 \quad \Rightarrow \text{ optimal!}$$

- Thus, the optimal solution is

$$x_1^* = 0.5$$
$$x_2^* = 0$$
$$x_3^* = 0$$
$$x_4^* = 0.125$$

giving an optimal cost $f^* = 1.0$

4.11: Quadratic Programming

- Quadratic programming is an approach in which the objective function is *quadratic* and the constraint functions $c_i(x)$ are *linear.*
- The general problem statement is:

$$\min_{x} q(x) \equiv \frac{1}{2}x^T G u + g^T u$$

$$\text{subject to } a_i^T x = b_i, \; i \in E$$
$$a_i^T x \le b_i, \; i \in I$$

  - Here, we assume that a solution x^* exists.
  - And, if the Hessian matrix G is positive definite, then x^* is a unique minimizing solution.
- First, we'll develop the *equality* constrained case, then generalize to the *inequality* constrained case using an *active set* strategy.
- Quadratic programming is different from Linear Programming in that it is possible to have meaningful problems in which there are no inequality constraints (due to *curvature* of the objective function).

Equality Constrained Quadratic Programming

- Problem statement:

$$\min_{x} q(x) \equiv \frac{1}{2}x^T G x + g^T x$$

$$\text{subject to } A^T x = b$$

  - Assume there are $m \le n$ constraints, and that A has rank m.
- Solution involves using constraints to eliminate variables (as we've done previously).

– First, define

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \quad A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad \boldsymbol{g} = \begin{bmatrix} \boldsymbol{g}_1 \\ \boldsymbol{g}_2 \end{bmatrix}, \quad G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$$

– Then we can write the constraint equations as:

$$A_1^T \boldsymbol{x}_1 + A_2^T \boldsymbol{x}_2 = \boldsymbol{b}$$

which, solving for $\boldsymbol{x}_1$ gives

$$\boldsymbol{x}_1 = A_1^{-T}\left(\boldsymbol{b} - A_2^T \boldsymbol{x}_2\right)$$

– Now, substituting into $q(\boldsymbol{x})$ yields the equivalent unconstrained minimization problem:

$$\min_{\boldsymbol{x}_2} \psi(\boldsymbol{x}_2)$$

where

$$\begin{aligned} \psi(\boldsymbol{x}_2) = {} & \frac{1}{2}\boldsymbol{x}_2^T \left(G_{22} - G_{21}A_1^{-T}A_2^T - A_2A_1^{-1}G_{12} + A_2A_1^{-1}G_{11}A_1^{-T}A_2^T\right)\boldsymbol{x}_2 \\ & + \boldsymbol{x}_2^T\left(G_{21} - A_2A_1^{-1}G_{11}\right)A_1^{-T}\boldsymbol{b} + \frac{1}{2}\boldsymbol{b}^T A_1^{-1}G_{11}A_1^{-T}\boldsymbol{b} \\ & + \boldsymbol{x}_2^T\left(\boldsymbol{g}_2 - A_2A_1^{-1}\boldsymbol{g}_1\right) + \boldsymbol{g}_1^T A_1^{-T}\boldsymbol{b} \end{aligned}$$

- A unique minimizing solution exists if the Hessian in the quadratic term is positive definite.
- In that case, $\boldsymbol{x}_2^*$ is found by solving the linear system

$$\nabla\psi(\boldsymbol{x}_2) = \mathbf{0}$$

and $\boldsymbol{x}_1^*$ is found by substitution.

Example

- Consider the quadratic programming problem given by:

$$\begin{aligned}
\min_{x_1,\,x_2,\,x_3} \; q(x_1,\, x_2,\, x_3) &= x_1^2 + x_2^2 + x_3^2 \\
&= \frac{1}{2}\boldsymbol{x}^T G \boldsymbol{x} \\
\text{subject to } x_1 + 2x_2 - x_3 &= 4 \\
x_1 - x_2 + x_3 &= -2 \\
A^T \boldsymbol{x} &= \boldsymbol{b}
\end{aligned}$$

  - This corresponds to the general problem with:

$$G = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}; \quad \boldsymbol{g} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 1 & 2 & -1 \\ 1 & -1 & 1 \end{bmatrix}; \quad \boldsymbol{b} = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$$

- Partitioning, we write ⇒

$$A^T = \left[\begin{array}{cc|c} 1 & 2 & -1 \\ 1 & -1 & 2 \end{array}\right] = \left[\begin{array}{c|c} A_{12}^T & A_3^T \end{array}\right]$$

and

$$\boldsymbol{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ \hline x_3 \end{array}\right] = \begin{bmatrix} \boldsymbol{x}_{12} \\ x_3 \end{bmatrix}$$

- We can invoke constraint equations to express $\boldsymbol{x}_{12}$ in terms of element x_3 :

$$\begin{bmatrix} A_{12}^T & A_3^T \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{12} \\ x_3 \end{bmatrix} = \boldsymbol{b}$$

$$\begin{aligned} A_{12}^T \boldsymbol{x}_{12} + A_3^T x_3 &= \boldsymbol{b} \\ \boldsymbol{x}_{12} &= A_{12}^{-T}\left(\boldsymbol{b} - A_3^T x_3\right) \\ &= A_{12}^{-T}\boldsymbol{b} - A_{12}^T A_3^T x_3 \end{aligned}$$

which for the present example gives:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 1/3 \\ -2/3 \end{bmatrix} x_3$$

- Substituting back into $q(\boldsymbol{x})$, we have

$$\begin{aligned} q(x) &= \frac{1}{2}\begin{bmatrix} \boldsymbol{x}_{12}^T & x_3^T \end{bmatrix} \left[\begin{array}{c|c} G_{11} & G_{12} \\ \hline G_{21} & G_{22} \end{array}\right] \begin{bmatrix} \boldsymbol{x}_{12} \\ x_3 \end{bmatrix} \\ &= \frac{1}{2}\left[\boldsymbol{x}_{12}^T G_{11}\boldsymbol{x}_{12} + 2\boldsymbol{x}_{12}^T G_{12} x_3 + x_3 G_{22} x_3\right] \\ &= \frac{1}{2}\Big[\left(A_{12}^{-T}\boldsymbol{b} - A_{12}^T A_3^T x_3\right)^T G_{11}\left(A_{12}^{-T}\boldsymbol{b} - A_{12}^T A_3^T x_3\right) \\ &\qquad +2\left(A_{12}^{-T}\boldsymbol{b} - A_{12}^T A_3 x_3\right) G_{12} x_3 + x_3 G_{22} x_3\Big] \end{aligned}$$

- Substituting values for G, A, and b, we obtain a quadratic expression in element x_3 :

$$\psi(x_3) = 0.5556x_3^2 + 2.6667x_3 + 4$$

- Since the Hessian

$$\frac{\partial^2 \psi}{\partial x_3^2} = 3.1111 > 0$$

we conclude the minimizer is unique and found by setting

$$\nabla \psi = \frac{\partial \psi}{\partial x_3} = 0$$

- This yields the solution

$$x_3^* = -.8571$$

and back substitution yields

$$x_1^* = .2857 \quad x_2^* = 1.4286$$

4.12: Quadratic Programming: Method of Lagrange Multipliers

- A more general approach for solving quadratic programming problems is via the method of Lagrange multipliers as seen previously.
- Consider the augmented cost function:

$$\bar{L}(\boldsymbol{x},\,\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{g}^T\boldsymbol{x} + \boldsymbol{\lambda}^T\left(A^T\boldsymbol{x} - \boldsymbol{b}\right)$$

- The stationarity condition is given by the equations:

$$\frac{\partial \bar{L}}{\partial \boldsymbol{x}} = G\boldsymbol{x} + \boldsymbol{g} + A\boldsymbol{\lambda} = \boldsymbol{0}$$

$$\frac{\partial \bar{L}}{\partial \boldsymbol{\lambda}} = A^T\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}$$

- These equations can be rearranged in the form of a linear system to give,

$$\begin{bmatrix} G & A \\ A^T & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{bmatrix} = -\begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{b} \end{bmatrix}$$

  – The matrix on the left is called the *Lagrangian Matrix* and is symmetric but not positive definite.
  – Several analytical methods of solution make use of the inverse form of the Lagrangian matrix

- Solving directly we may write:

$$\boldsymbol{\lambda} = -\left(A^T G^{-1} A\right)\left(\boldsymbol{b} + A^T G^{-1}\boldsymbol{g}\right)$$

$$\boldsymbol{x} = -G^{-1}g - G^{-1}A\boldsymbol{\lambda}$$

- It is interesting to note that $\boldsymbol{x}$ can be written in the form of two terms:

$$\boldsymbol{x} = -G^{-1}\boldsymbol{g} - G^{-1}A\,\boldsymbol{\lambda} = \boldsymbol{x}^0 - G^{-1}A\boldsymbol{\lambda}$$

where the first term $\boldsymbol{x}^0$ is the global minimum solution to the unconstrained problem and the second term is a correction due to the equality constraints.

Example

- Returning to our previous example, the solution without constraints is given by:

$$\boldsymbol{x}^0 = -G^{-1}\boldsymbol{g} = \boldsymbol{0}$$

- Solving for the Lagrange multipliers,

$$\boldsymbol{\lambda} = -\left(\begin{bmatrix} 1 & 2 & -1 \\ 1 & -1 & 1 \end{bmatrix}\begin{bmatrix} 0.5 & & \\ & 0.5 & \\ & & 0.5 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 2 & -1 \\ -1 & 1 \end{bmatrix}\right)^{-1}$$

$$\times\left(\begin{bmatrix} 4 \\ -2 \end{bmatrix} + \begin{bmatrix} 1 & 2 & -1 \\ 1 & -1 & 1 \end{bmatrix}\begin{bmatrix} 0.5 & & \\ & 0.5 & \\ & & 0.5 \end{bmatrix}\left[\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right]\right)$$

giving

$$\boldsymbol{\lambda} = \begin{bmatrix} -1.1429 \\ 0.5714 \end{bmatrix}$$

- The minimizing solution vector is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.5 & & \\ & 0.5 & \\ & & 0.5 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 2 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1.1429 \\ -.5714 \end{bmatrix} = \begin{bmatrix} 0.2857 \\ 1.4286 \\ -.8571 \end{bmatrix}$$

4.13: Quadratic Programming: Inequality Constraints

- Inequality constrained problems include constraints from the set $i \in I$ where the number of inequality constraints could be larger than the number of decision variables.
- The constraint set, $A^T\boldsymbol{x} \le \boldsymbol{b}$ may include both active and inactive constraints where a constraint $\boldsymbol{a}_i^T\boldsymbol{x} \le b_i$ is said to be *active* if $\boldsymbol{a}_i^T\boldsymbol{x} = b_i$ and *inactive* if $\boldsymbol{a}_i^T\boldsymbol{x} < b_i$.

Kuhn-Tucker Conditions

- The necessary conditions for satisfaction of this optimization problem are given by the KUHN-TUCKER conditions:

$$\begin{aligned} G\boldsymbol{x} + \boldsymbol{g} + A\boldsymbol{\lambda} &= \boldsymbol{0} \\ A^T\boldsymbol{x} - \boldsymbol{b} &\le \boldsymbol{0} \\ \boldsymbol{\lambda}^T\left(A^T\boldsymbol{x} - \boldsymbol{b}\right) &= 0 \\ \boldsymbol{\lambda} &\ge \boldsymbol{0} \end{aligned}$$

- We can express the Kuhn-Tucker conditions in terms of the *active* constraints as:

$$\begin{aligned} G\boldsymbol{x} + \boldsymbol{g} + \sum_{i \in \mathscr{A}} \lambda_i \boldsymbol{a}_i &= 0 \\ \boldsymbol{a}_i^T x - b_i &= 0 \quad i \in \mathscr{A} \\ \boldsymbol{a}_i^T x - b_i &< 0 \quad i \notin \mathscr{A} \\ \lambda_i &\ge 0 \quad i \in \mathscr{A} \\ \lambda_i &= 0 \quad i \notin \mathscr{A} \end{aligned}$$

  - In other words, the active constraints are equality constraints.

- Assuming that $A_{\mathscr{A}}^T$ and $\boldsymbol{\lambda}_{\mathscr{A}}$ are known (where $\mathscr{A}$ denotes the active set), the original problem can be replaced by the corresponding problem having only equality constraints:

$$\boldsymbol{\lambda}_{\mathscr{A}} = -\left(A_{\mathscr{A}}^T G^{-1} A_{\mathscr{A}}\right)^{-1}\left(\boldsymbol{b}_{\mathscr{A}} + A_{\mathscr{A}}^T G^{-1}\boldsymbol{g}\right)$$
$$\boldsymbol{x} = -G^{-1}\left(\boldsymbol{g} + A_{\mathscr{A}}\boldsymbol{\lambda}_{\mathscr{A}}\right)$$

Active Set Methods

- Active set methods take advantage of the solution to equality constraint problems in order to solve more general inequality contraint problems.
- BASIC IDEA: Define at each algorithm step a set of constraints (called the *working set*) that is treated as the active set.
  - The working set, $\mathscr{W}$, is a subset of the active set $\mathscr{A}$ at the current point; the vectors $\boldsymbol{a}_i \in \mathscr{W}$ are linearly independent.
  - The current point is *feasible* for the working set.
  - The algorithm proceeds to an improved point.
  - An equality constrained problem is solved at each step.
- If all $\lambda_i \geq 0$, the point is a local solution .
- If some $\lambda_i < 0$, then the objective function can be decreased further by relaxing the corresponding constraint.

Example [Wang, pg. 60]

- Here we develop a solution to the following problem:

$$\min_{\boldsymbol{x}} q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} -2 & -3 & -1 \end{bmatrix} \boldsymbol{x}$$

$$\begin{aligned} \text{subject to:} \quad x_1 + x_2 + x_3 &\le 1 \\ 3x_1 - 2x_2 - 3x_3 &\le 1 \\ x_1 - 3x_2 + 2x_3 &\le 1 \end{aligned}$$

- Relevent matrices are:

$$G = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}; \quad \boldsymbol{g} = \begin{bmatrix} -2 \\ -3 \\ -1 \end{bmatrix}; \quad A^T = \begin{bmatrix} 1 & 1 & 1 \\ 3 & -2 & -3 \\ 1 & -3 & 2 \end{bmatrix}; \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

- A feasible solution of the equality constraints exists since the linear equations $A^T\boldsymbol{x} = \boldsymbol{b}$ are well determined (i.e., A^T is full rank)
- Using the three equality constraints as the first working set, we calculate

$$\boldsymbol{\lambda} = -\left(A^T G^{-1} A\right)^{-1}\left(\boldsymbol{b} + A^T G^{-1}\boldsymbol{g}\right) = \begin{bmatrix} 1.6873 \\ 0.0309 \\ -0.4352 \end{bmatrix}$$

- Since $\lambda_3 < 0$, we conclude the third constraint equation is inactive, and omit it from the active set, $\mathscr{A}$
- We then solve the reduced problem,

$$\min_{\boldsymbol{x}} q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} -2 & -3 & -1 \end{bmatrix} \boldsymbol{x}$$

$$\begin{aligned} \text{subject to}: \quad x_1 + x_2 + x_3 &\le 1 \\ 3x_1 - 2x_2 - 3x_3 &\le 1 \end{aligned}$$

- Solving now for the remaining Lagrange multipliers,

$$\boldsymbol{\lambda} = \begin{bmatrix} 1.6452 \\ -.0323 \end{bmatrix}$$

  - As before, we see that $\lambda_2 < 0$, so conclude that the second constraint equation is inactive, and omit it from $\mathscr{A}$.

- We are now left with an equality constraint problem having just the single constraint,

$$x_1 + x_2 + x_3 = 1$$

- Solving this problem, we obtain

$$\lambda = 1.6$$

  and compute

$$\boldsymbol{x}^* = \begin{bmatrix} 0.3333 \\ 1.3333 \\ -0.6667 \end{bmatrix}$$

4.14: Primal-Dual Method

- Active methods are a subset of the *primal* methods, wherein solutions are based directly on the decision (i.e., *primal*) variables.
- Computationally, this method can become burdensome if the number of contraints is large.
- A *dual* method can often be used to reach the solution of a primal method while realizing a computational savings.
- For our present problem, we will identify the Lagrange multipliers as the dual variables; we derive the dual problem as follows:
  - Assuming feasibility, the primal problem is equivalent to:

$$\max_{\boldsymbol{\lambda}\geq\mathbf{0}}\min_{\boldsymbol{x}}\left[\frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x}+\boldsymbol{g}^T\boldsymbol{x}+\boldsymbol{\lambda}^T\left(A^T\boldsymbol{x}-\boldsymbol{b}\right)\right]$$

  - The minimum over $\boldsymbol{x}$ is *unconstrained* and given by

$$\boldsymbol{x}=-G^{-1}\left(\boldsymbol{g}+A\boldsymbol{\lambda}\right)$$

  - Substituting into the above expression, we write the dual problem as:

$$\max_{\boldsymbol{\lambda}\geq\mathbf{0}}\left(-\frac{1}{2}\boldsymbol{\lambda}^T H\boldsymbol{\lambda}-\boldsymbol{\lambda}^T K-\frac{1}{2}\boldsymbol{g}^T G^{-1}\boldsymbol{g}\right)$$

    where

$$H=A^T G^{-1}A$$
$$K=\boldsymbol{b}+A^T G^{-1}\boldsymbol{g}$$

  - This is now equivalent to the quadratic programming problem:

$$\min_{\boldsymbol{\lambda}\geq\mathbf{0}}\left(\frac{1}{2}\boldsymbol{\lambda}^T H\boldsymbol{\lambda}+\boldsymbol{\lambda}^T K+\frac{1}{2}\boldsymbol{g}^T G^{-1}\boldsymbol{g}\right)$$

  - Note that this form of the problem may be easier to solve than the primal problem since the constraints are simpler.

- A full derivation of the dual form of the quadratic programming problem appears in Appendix 4.1.

4.15: Hildreth's Quadratic Programming Algorithm

- Hildreth's procedure is a robust and systematic way to solve the dual quadratic programming problem.
- The basic idea is to vary the Lagrange multiplier vector elements, λ_i, one at a time, and adjust each element such as to minimize the objective function.
- A single iteration through the cycle may be expresses as:

$$\lambda_i^{(k+1)} = \max\left(0,\ w_i^{(k+1)}\right)$$

where

$$w_i^{(k+1)} = -\frac{1}{h_{ii}}\left[k_i + \sum_{j=1}^{i-1} h_{ij}\lambda_j^{(k+1)} + \sum_{j=i+1}^{n} h_{ij}\lambda_j^{(m)}\right]$$

  – Here, we define h_{ij} to be the ij-th element of the matrix H and k_i is the i-th element of the vector K.

- The Hildreth algorithm implements an iterative solution of the linear equations:

$$\begin{aligned}\boldsymbol{\lambda} &= -\left(A^T G^{-1} A\right)^{-1}\left(\boldsymbol{b} + A^T G^{-1}\boldsymbol{g}\right)\\ &= -H^{-1}K\end{aligned}$$

which can be equivalently expressed as:

$$H\boldsymbol{\lambda} = -K$$

- Note that this approach avoids the need to perform a matrix inverse, which leads to a robust algorithm.
- Once the vector $\boldsymbol{\lambda}$ converges to $\boldsymbol{\lambda}^*$, the solution vector is found as:

$$\boldsymbol{x}^* = -G^{-1}\left(\boldsymbol{g} + A\boldsymbol{\lambda}^*\right)$$

- A proof of Hildreth algorithm convergence is presented in Appendix 4.2.

Example [Wang, pg. 64]

- Consider the optimization problem

$$\min_{\boldsymbol{x}} q(\boldsymbol{x}) = x_1^2 - x_1 x_2 + \frac{1}{2}x_2^2 - x_1$$

$$\begin{aligned} \text{subject to } 3x_1 + 2x_2 &\le 4 \\ x_1 \quad\quad &\ge 0 \\ x_2 \quad\quad &\ge 0 \end{aligned}$$

- In standard form it can be shown that

$$G = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}; \quad \boldsymbol{g} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}; \quad A^T = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 3 & 2 \end{bmatrix}; \quad \boldsymbol{b} = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix}$$

- The global optimum (unconstrained) minimum is:

$$\boldsymbol{x}^0 = -G^{-1}\boldsymbol{g} = -\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- It is easily seen from the figure that the optimum solution violates the inequality constraints (is not within the feasible region).

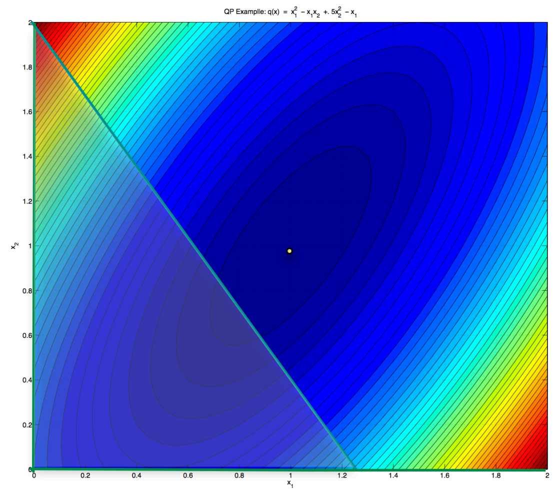

Figure 4.10: Quadratic Programming Example: Feasible Region

- In order to implement the Hildreth algorithm, we form the matrices H and K:

$$H = A^T G^{-1} A = \begin{bmatrix} 1 & 1 & -5 \\ 1 & 2 & -7 \\ -5 & -7 & 29 \end{bmatrix}; \quad K = \boldsymbol{b} + A^T G^{-1} \boldsymbol{g} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

- Iteration $k = 0$
  - $\lambda_1^{(0)} = \lambda_2^{(0)} = \lambda_3^{(0)} = 0$
- Iteration $k = 1$

$$\begin{aligned} w_1^{(1)} + 1 &= 0 \\ \lambda_1^{(1)} + 2w_2^{(1)} + 1 &= 0 \\ -5\lambda_1^{(1)} - 7\lambda_2^{(1)} + 29w_3^{(1)} - 1 &= 0 \end{aligned}$$

  - solving gives:

$$\lambda_1^{(1)} = \max\left(0,\ w_1^{(1)}\right) = 0$$

$$\lambda_2^{(1)} = \max\left(0,\ w_2^{(1)}\right) = 0$$

$$\lambda_3^{(1)} = \max\left(0,\ w_3^{(1)}\right) = .0345$$

- Iteration $k = 2$

$$w_1^{(2)} + \lambda_2^{(1)} - 5\lambda_3^{(1)} + 1 \quad = \quad 0$$

$$\lambda_1^{(2)} + 2w_2^{(2)} - 7\lambda_3^{(1)} + 1 = 0$$

$$-5\lambda_1^{(2)} - 7\lambda_2^{(2)} + 29w_3^{(2)} - 1 = 0$$

  – solving gives:

$$\lambda_1^{(2)} = \max\left(0,\ w_1^{(2)}\right) = 0$$

$$\lambda_2^{(2)} = \max\left(0,\ w_2^{(2)}\right) = 0$$

$$\lambda_3^{(2)} = \max\left(0,\ w_3^{(2)}\right) = .0345$$

  – thus, the iterative process has converged

- The optimal solution is given by:

$$x^* = -G^{-1}\left(\boldsymbol{g} + A\boldsymbol{\lambda}^*\right)$$

$$= -G^{-1}\boldsymbol{g} - G^{-1}A\boldsymbol{\lambda}^* = \boldsymbol{x}^0 - G^{-1}A\boldsymbol{\lambda}^*$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -1 & 0 & 3 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0.0345 \end{bmatrix} = \begin{bmatrix} 0.8276 \\ 0.7586 \end{bmatrix}$$

Appendix 4.A

Derivation of the Dual Optimization Problem

- Starting with the primal problem,

$$L^* = \max_{\boldsymbol{\lambda} \geq 0} \min_{\boldsymbol{x}} \left[\frac{1}{2} \boldsymbol{x}^T G \boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{g} + \boldsymbol{\lambda}^T (A^T \boldsymbol{x} - \boldsymbol{b}) \right]$$

we solve for the unconstrained minimizer

$$\boldsymbol{x}^o = -G^{-1}(\boldsymbol{g} + A\boldsymbol{\lambda})$$

- Subsituting $\boldsymbol{x}^o$ into the primal problem, we obtain

$$L^* = \max_{\boldsymbol{\lambda} \geq 0} \left\{ \frac{1}{2} (\boldsymbol{g} + A\boldsymbol{\lambda})^T G^{-1} G G^{-1} (\boldsymbol{g} + A\boldsymbol{\lambda}) \right.$$

$$\left. - \, (\boldsymbol{g} + A\boldsymbol{\lambda})^T G^{-1} \boldsymbol{g} + \boldsymbol{\lambda}^T \left(A^T \left(-G^{-1}(\boldsymbol{g} + A\boldsymbol{\lambda}) \right) - \boldsymbol{b} \right) \right\}$$

- Expanding and cancelling terms,

$$= \max_{\boldsymbol{\lambda} \geq 0} \left\{ \frac{1}{2} \left(\boldsymbol{\lambda}^T A^T + \boldsymbol{g}^T \right) G^{-1} (\boldsymbol{g} + A\boldsymbol{\lambda}) \right.$$

$$\left. - \left(\boldsymbol{\lambda}^T A^T + \boldsymbol{g}^T \right) G^{-1} \boldsymbol{g} + \boldsymbol{\lambda}^T \left[-A^T G^{-1} (\boldsymbol{g} + A\boldsymbol{\lambda}) - \boldsymbol{b} \right] \right\}$$

$$= \max_{\boldsymbol{\lambda} \geq 0} \left\{ \frac{1}{2} \left[\boldsymbol{\lambda}^T A^T G^{-1} (\boldsymbol{g} + A\boldsymbol{\lambda}) + \boldsymbol{g}^T G^{-1} (\boldsymbol{g} + A\boldsymbol{\lambda}) \right] \right.$$

$$\left. - \boldsymbol{\lambda}^T A^T G^{-1} \boldsymbol{g} - \boldsymbol{g}^T G^{-1} \boldsymbol{g} + \boldsymbol{\lambda}^T \left[-A^T G^{-1} \boldsymbol{g} - A^T G^{-1} A \boldsymbol{\lambda} - \boldsymbol{b} \right] \right\}$$

$$= \max_{\boldsymbol{\lambda} \geq 0} \left\{ \frac{1}{2} \boldsymbol{\lambda}^T A^T G^{-1} \boldsymbol{g} + \frac{1}{2} \boldsymbol{\lambda}^T A^T G^{-1} A \boldsymbol{\lambda} \right.$$

$$+ \frac{1}{2} \boldsymbol{g}^T G^{-1} \boldsymbol{g} + \frac{1}{2} \boldsymbol{g}^T G^{-1} A \boldsymbol{\lambda} - \boldsymbol{\lambda}^T A^T G^{-1} \boldsymbol{g} - \boldsymbol{g}^T G^{-1} \boldsymbol{g}$$

$$\left. - \boldsymbol{\lambda}^T A^T G^{-1} \boldsymbol{g} - \boldsymbol{\lambda}^T A^T G^{-1} A \boldsymbol{\lambda} - \boldsymbol{\lambda}^T \boldsymbol{b} \right\}$$

- And finally,

$$L^* = \max_{\boldsymbol{\lambda} \geq 0} \left\{ -\boldsymbol{\lambda}^T A^T G^{-1} \boldsymbol{g} - \frac{1}{2} \boldsymbol{\lambda}^T A^T G^{-1} A \boldsymbol{\lambda} - \frac{1}{2} \boldsymbol{g}^T G^{-1} \boldsymbol{g} - \boldsymbol{\lambda}^T \boldsymbol{b} \right\}$$

giving the result.

Appendix 4.B

Proof of Convergence for the Hildreth Procedure

- We assert here that if matrix $P = A^T G^{-1} A$ is positive definite in the dual optimization problem given by

$$\min_{\boldsymbol{\lambda} \geq 0} L = \left(\frac{1}{2} \boldsymbol{\lambda}^T P \boldsymbol{\lambda} + \boldsymbol{\lambda}^T K + \frac{1}{2} \boldsymbol{g}^T G^{-1} \boldsymbol{g} \right),$$

  where

$$K = \boldsymbol{b} + A^T G^{-1} \boldsymbol{g},$$

  then the sequence

$$\left\{ \boldsymbol{\lambda}^{(m)} \right\} \rightarrow \boldsymbol{\lambda}^*.$$

- Let us define by $\mathcal{H}\left(\boldsymbol{\lambda}^{(m)}\right)$ the Hildreth operator such that

$$\boldsymbol{\lambda}^{(m+1)} = \mathcal{H}\left(\boldsymbol{\lambda}^{(m)}\right).$$

- We shall first show that

$$\mathcal{H}(\boldsymbol{\lambda}) = \boldsymbol{\lambda} \quad \Rightarrow \quad \boldsymbol{\lambda} = \boldsymbol{\lambda}^*$$

- First take any $\boldsymbol{\lambda} \geq \mathbf{0}$, $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}^*$, then

$$L(\boldsymbol{\lambda}^*) - J(\boldsymbol{\lambda}) < 0.$$

- Now let $\boldsymbol{\lambda}^* = \boldsymbol{\lambda} + \boldsymbol{\delta}$, we can now write

$$\begin{aligned} L(\boldsymbol{\lambda}^*) &= \frac{1}{2}(\boldsymbol{\lambda} + \boldsymbol{\delta})^T P (\boldsymbol{\lambda} + \boldsymbol{\delta}) + (\boldsymbol{\lambda} + \boldsymbol{\delta})^T \boldsymbol{K} + \frac{1}{2} \boldsymbol{g}^T G^{-1} \boldsymbol{g} \\ &= \frac{1}{2}\left(\boldsymbol{\lambda}^T P \boldsymbol{\lambda} + 2\boldsymbol{\lambda}^T \boldsymbol{\delta} + \boldsymbol{\delta}^T P \boldsymbol{\delta}\right) + \boldsymbol{\lambda}^T \boldsymbol{K} + \boldsymbol{\delta}^T \boldsymbol{K} + \frac{1}{2} \boldsymbol{g}^T G^{-1} \boldsymbol{g} \\ &= \left(\frac{1}{2} \boldsymbol{\lambda}^T P \boldsymbol{\lambda} + \boldsymbol{\lambda}^T \boldsymbol{K} + \frac{1}{2} \boldsymbol{g}^T G^{-1} \boldsymbol{g} \right) + \boldsymbol{\lambda}^T P \boldsymbol{\delta} + \boldsymbol{\delta}^T \boldsymbol{K} + \frac{1}{2} \boldsymbol{\delta}^T P \boldsymbol{\delta} \\ &= L(\boldsymbol{\lambda}) + \boldsymbol{\lambda}^T P \boldsymbol{\delta} + \boldsymbol{\delta}^T \boldsymbol{K} + \frac{1}{2} \boldsymbol{\delta}^T P \boldsymbol{\delta} \end{aligned}$$

- Thus we have,

$$L\left(\boldsymbol{\lambda}^{*}\right)-L(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^{T} P \boldsymbol{\delta}+\boldsymbol{\delta}^{T} \boldsymbol{K}+\frac{1}{2} \boldsymbol{\delta}^{T} P \boldsymbol{\delta}$$
$$= \left(\frac{\partial L}{\partial \boldsymbol{\lambda}}\right) \boldsymbol{\delta}+\frac{1}{2} \boldsymbol{\delta}^{T} P \boldsymbol{\delta}$$

- Since the term $\frac{1}{2} \boldsymbol{\delta}^{T} P \boldsymbol{\delta}$ must be positive, it is clear that $\left(\frac{\partial L}{\partial \boldsymbol{\lambda}}\right) \boldsymbol{\delta}$ must be negative, since $L\left(\boldsymbol{\lambda}^{*}\right)-L(\boldsymbol{\lambda})$ is negative.
- Hence, $L(\boldsymbol{\lambda})$ can be reduced by changing one element of $\boldsymbol{\lambda}$, and the operator $\mathcal{H}$ will do this.

- Next, we note that the sequence $\{\boldsymbol{\lambda}^{(m)}\}$ is contained in the bounded set $\{\boldsymbol{\lambda} \,|L(\boldsymbol{\lambda}) \le L(\boldsymbol{\lambda}^{(0)})\}$ and therefore has a limit.
- Let $\boldsymbol{\lambda}^{\infty}$ be a limit and let $\{\boldsymbol{\lambda}^{[r]}\}$ be a sub-sequence that approaches $\boldsymbol{\lambda}^{\infty}$; we now attempt to show that the conjecture $\boldsymbol{\lambda}^{\infty} \neq \boldsymbol{\lambda}^{*}$ will lead to a contradiction.
  - If$\boldsymbol{\lambda}^{\infty} \neq \boldsymbol{\lambda}^{*}$, then
  $$L\left(\boldsymbol{\lambda}^{\infty}\right)-L\left(\mathcal{H}\left(\boldsymbol{\lambda}^{\infty}\right)\right)=\epsilon>0.$$
  - Continuity of $L(\lambda)$ in the bounded region containing $\{\boldsymbol{\lambda}^{(m)}\}$ assures that for any such ϵ there exsits a $\delta>0$ such that
  $$\left\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^{*}\right\|<\delta \quad \Rightarrow \quad\left|L(\boldsymbol{\lambda})-L\left(\boldsymbol{\lambda}^{*}\right)\right|<\epsilon$$
  - From the continuity of the operator $\mathcal{H}$, there exists a positive integer, R, such that
  $$r>R \quad \rightarrow \quad \left\|\boldsymbol{\lambda}^{[r]}-\boldsymbol{\lambda}^{\infty}\right\|<\rho.$$
  - For such an r, say r^{*}, let p^{*} be the index of $\boldsymbol{\lambda}^{[r^{*}]}$ in the original sequence; for such a p^{*},
  $$\left|L\left(\boldsymbol{\lambda}^{(p^{*}+1)}\right)-L\left(\mathcal{H}\left(\boldsymbol{\lambda}^{\infty}\right)\right)\right|<\epsilon$$

and therefore,

$$L\left(\boldsymbol{\lambda}^{(p^*+1)}\right) < L\left(\boldsymbol{\lambda}^{\infty}\right)$$

which contradicts the definition of $\left\{\boldsymbol{\lambda}^{(m)}\right\}$.